The Analysis of A Contingency Table

Ludovic Lebart

Descriptions of brand-switching matrices

According to the views of this contributor, it is meaningful to use exploratory techniques such as "principal axes techniques", (namely Principal component analysis or Correspondence analysis and their variants), mostly in two cases:

1. The data set is very large, and the a priori knowledge about is content is very poor: a powerful purely a descriptive tool is then needed
2. The data set comprises several subsets corresponding to different themes, some of them having an explanatory role, for example, a set of variables describing car-purchase behavior on the one hand and a set of socioeconomic characteristics (or lifestyle variables) on the other. The positioning of these latter variables onto the typologies constructed using the first set is certainly a very rewarding process in terms of cost, time, and power of investigation.

Obviously, the proposed example does not belong to these categories. Thus, it will be considered here an (interesting) academic exercise.

Two techniques are used to perform a multivariate descriptive analysis of the data matrix: Correspondence Analysis (CA) mainly used here as an ordination technique, and hierarchical classification (HC). We proceed in two steps:

- Overall description of the "behaviours" on the 15 makes using both CA and HC, leading to two elementary (but suggestive) printouts of the initial data matrix.
- Focussing on the subpopulation of switchers (the diagonal elements of the matrix, representing the "loyal clients", being replaced by zero values), a new description through CA emphasizes the non-symmetrical aspects of the relationships between *previous makes* and *new makes*.

Note that the data matrix belongs to the family of "confusion matrices", whose rows often represent some sent out stimuli (colours for example) and columns represent the recognized stimuli. Such matrices are almost symmetric. They are generally provided with a heavily loaded diagonal (more distinct are the stimuli, more loaded is the diagonal).

Description of the global matrix (French data, 198D)

Hierarchical classification (using Chi-squared distances on the matrix with a reduced[1] diagonal, and Ward's criterion, see Figure 1) put forward two dominant groups: The luxury makes on the one hand, the more usual cars on the other, this latter group being split into two subgroups: French makes and other European makes.

Table 1 summarizes the row-profiles and column-profiles along the first axis of CA (these profiles, expressed as percentages, are computed using the actual sum of the rows and of the columns, disregarding the totals provided separately). It is obvious, looking at the area with

[1] The unloading of the diagonal is an usual procedure in the analysis of confusion matrices. It leads to more meaningful measures of similarity (see for instance Benzecri, "*Sur l'analyse des matrices de confusion", Revue de statistique appliquee*, 18, p 5-62, 1970). The applied reduction rate in this case is 35%. It corresponds to the following model: 35% of the owners are systematically "loyal", whatever the make, and can be removed without damage of the analysis (only "seat," a recent make in the French market, has a lower percentage of loyalists.

almost empty cells in the upper right part of Table 1.A, that those who leave *Lada, Citroen, Renault, Peugeot, Fiat, G.M., Seat, etc*. will very seldom choose *Alfa, Volvo, BMW or Mercedes*. Similarly, those who now purchase the makes of the former list are not previous owners of the second list cars (lower-left part of Table 1.B). This clear-cut feature accounts for the main two subtrees of Figure 1.

These two modes of representation, although elementary, can provide much more information. Figure 1 shows that among the French makes, *Citroen* has a specific behaviour (refer to the so-called *citroenists* in this country), as well as *Lada* (the only eastern make) among the "ordinary" cars and *Alfa* among the luxury cars. Most diagonal elements of Table 1.A contain the percentages of loyal owners (not all elements, since the new orders of rows and columns are not quite identical). In decreasing order: *Mercedes* (69.7%), *Renault, Peugeot, VW, Citroen, Ford, G.M., Saab, Volvo, Fiat, Alfa, Lada, Rover, Seat* (26.3%).

The percentages of *loyal buyers* can be read in most diagonal elements of Table 1.B. For some makes, the *expanding makes*, this latter percentage is much smaller (*Mercedes: 52%* instead of 70%; *Saab*: 34% instead of 52%, *Seat*: 11% instead of 26%. For some others, it is larger (*Renault:* 72% instead of 63%, *Citroen* 67% instead of 55%, *Alfa*: 56% instead of 37%...)

The Subpopulation of Switchers

If we remove completely the diagonal elements of the initial matrix, we obtain through CA the so-called "inverse factors", clearly separating the owners (lowercase italic labels in Figure 2) and the purchasers (circled uppercase labels in Figure 2). The patterns of owner-points and purchaser-points are quite different, since the corresponding profiles are obviously very different (see Figure 2). The arrows point out some of the more noticeable links ("departures" from one make or "arrivals" into another make).

Note that the three (underlined) French makes, representing about 60% of the total market, play a prominent role, since they "feed" the newcomers. They are located on the periphery of the display. The dotted line joining the owners of *Peugeot* (left) and of *Renault* (right) to *Citroen* stresses the relative weakness of this link. We read for example in Table 2A (keeping in mind that these new percentages are computed within the set of switchers) that, among the switchers, 37% of *Peugeot* owners go to *Renault*, whereas only 14% changes for *Citroen*.

Similarly, 32% leave *Renault* for *Peugeot*, vs. 13% for *Citroen*. On the contrary, those who leave *Citroen* (lower part of Figure 2, and 5th row of Table 2.A) join massively the two other French makes. These three French makes (especially *Renault* and *Peugeot*, see the weights of the three corresponding rows in Table 2.B) provide a majority of clients to all makes (up to 73% of the clients for *Lada*), if we except *Saab* whose figures are not significant. Note the location of those who leave *VW* join *Peugeot*, the largest rate for a foreign car joining a French car (a similar phenomenon is observed for *Lada*). Note also that the point representing the profile of *VW* purchasers is located much higher in the area of luxury cars. Some figures of table 2 account for this result: 21% of those who leave *Volvo* and 22% of those who leave *Saab* join *VW*.

A somewhat surprising result is suggested by the location of the owner-points of *Volvo* and *Mercedes* in the left part of Figure 2: those who leave *Volvo* and *Mercedes* choose *Renault* twice as frequently than *Peugeot* (at the time of this survey, the top model of *Peugeot* was not yet on sale... such results stressing the limitation of too global statistical data).

Among the significant links defining the cluster of luxury cars, we must note that 29% of *Mercedes* owners choose *BMW*, the highest rate of the whole table 2.A for a foreign car. We note also that *Fiat* provides 14% of the new purchaser of *Alfa*, by far the highest foreign percentage for this make.

As a conclusion...

Three national makes cover about two third of the market, but concern also two third of the switchers.

The *opening of the French market* is mainly due to *Renault* owners (55% of them purchase a foreign car, 49% for *Peugeot*, 35% for *Citroen*).

A *cluster of luxury makes* is clearly defined, but the luxury models of other brands are artificially excluded, since they are merged with smaller models. The figures relating to makes with large ranges of models have little significance, due to the heterogeneity of the cars they represent. In some respects, it would be more accurate to speak of "specialized makes without inexpensive models" instead of "luxury makes". Nevertheless, one can observe:

> A *status effect*, well described by both figure 1 and Table 1, with the restriction of interpretation mentioned above.
>
> Probably a *country preference effect*, ("cross loyalty" within French, Swedish, German, Italian makes). This effect is partially visible in Figure 1, and is reinforced by many of the above observations.

These two effects are not independent, since there is an obvious link between the status and the country of origin.

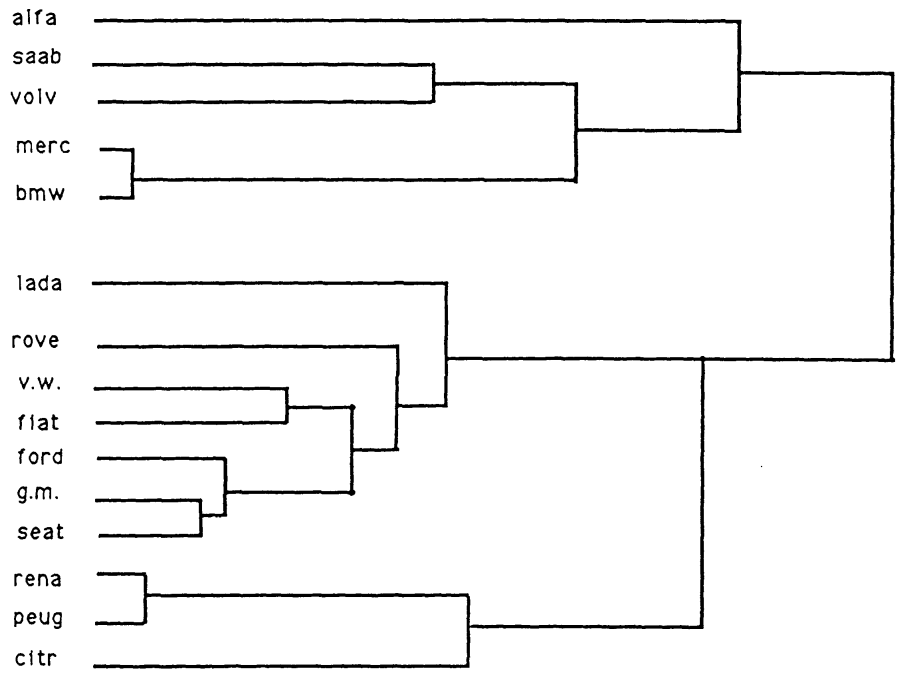

FIGURE 1

Three main clusters of makes issued from a hierarchical classification of profiles

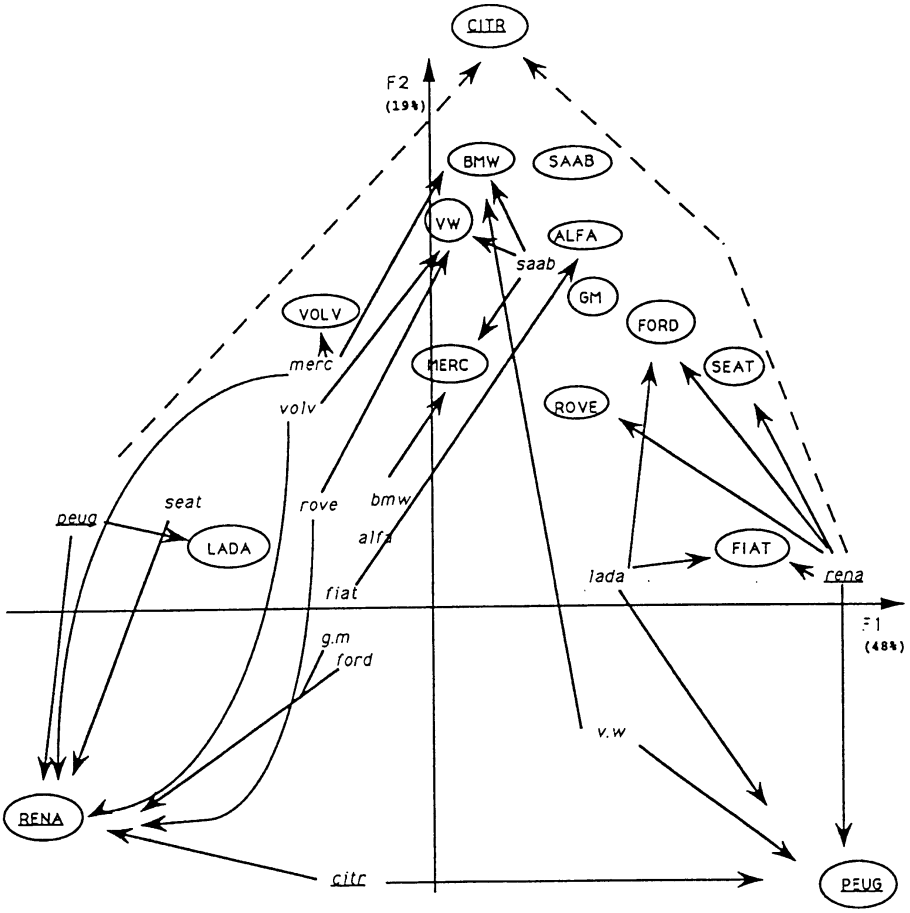

FIGURE 2

CHANGES OF MAKES: CORRESPONDENCE ANALYSIS RESTRICTED TO ACTUAL SWITCHERS (NULL DIAGONAL ELEMENTS IN THE MATRIX)

Table 1

Reordering of Rows and Columns According to the First Axis of Correspondence Analysis

1.A Row Profiles (The sum of the elements of each row equals 100)

	LADA	CIT	REN	PEUG	FIAT	GM	SEAT	FORD	ROV	VW	ALFA	VOLV	BMW	SAAB	MERC
Lada	39.6	7.0	9.4	14.6	7.6	7.0	1.8	6.4	2.3	4.1	.0	.0	.0	.0	.0
Cit	.6	55.5	14.6	14.2	4.2	2.1	.9	3.0	.9	2.8	.2	.3	.4	.0	.5
Ren	.5	4.6	63.5	11.7	4.4	3.3	1.4	4.0	1.0	4.0	.3	.3	.5	.1	.4
Peug	.7	5.6	14.9	60.0	4.4	3.4	1.2	4.0	1.0	4.1	.3	.4	.5	.0	.5
Seat	.0	9.5	21.2	7.3	7.3	8.0	26.3	7.3	2.2	9.5	1.5	.0	.0	.0	.0
Fiat	1.1	6.1	13.0	11.8	46.2	4.3	1.5	4.4	2.4	7.2	1.0	.4	.4	.1	.3
GM	1.0	2.9	11.7	9.9	3.4	52.8	1.8	7.3	1.0	5.7	.3	.6	1.0	.0	.6
Ford	.6	3.5	11.4	9.5	4.1	5.9	1.8	54.5	1.6	4.7	.3	.5	.8	.1	.6
Rov	1.8	5.5	13.8	9.2	7.6	.9	1.8	9.2	35.2	12.2	.6	.6	1.2	.0	.3
vw	.1	3.6	8.3	12.8	3.7	4.3	1.8	4.7	.9	56.0	.4	.7	1.5	.2	.7
Alfa	.0	7.3	13.5	10.4	7.3	2.7	2.3	5.4	1.9	6.6	37.5	1.5	1.9	.4	1.2
Volv	.0	6.7	10.4	5.5	3.0	3.7	2.4	2.4	1.8	11.0	.6	47.6	1.8	1.2	1.8
BMW	.0	5.6	11.1	10.8	2.5	1.7	1.1	3.9	1.7	8.1	1.1	1.1	45.3	.8	5.3
Saab	.0	.0	.0	.0	5.3	.0	.0	.0	.0	10.5	5.3	5.3	10.5	52.6	10.5
Merc	.0	1.5	5.1	2.1	1.5	1.0	.5	2.6	1.5	2.6	1.0	1.5	8.7	.5	69.7

1.B Column Profiles (The sum of the elements of each column equals 100

	LADA	CIT	REN	PEUG	FIAT	GM	SEAT	FORD	ROV	VW	ALFA	VOLV	BMW	SAAB	MERC
Lada	34.3	.4	.2	.5	.9	1.1	.9	.7	1.1	.4	.0	.0	.0	.0	.0
Cit	10.6	66.7	7.1	9.4	9.9	6.2	8.5	6.3	8.1	5.3	3.5	5.8	4.2	3.4	5.8
Ren	20.2	13.0	72.0	18.2	24.3	23.5	32.0	19.9	21.7	18.4	12.7	14.0	11.9	13.8	12.7
Peug	17.2	10.0	10.8	59.4	12.0	15.5	18.1	12.6	13.1	11.8	7.5	11.7	7.4	6.9	8.8
Seat	.0	.5	.4	.2	.7	1.0	10.9	.6	.8	.8	1.2	.0	.0	.0	.0
Fiat	6.1	2.5	2.2	2.7	38.3	4.5	5.1	3.2	7.5	4.9	6.4	2.9	1.6	3.4	1.2
GM	3.5	.7	1.2	1.4	1.7	33.5	3.6	3.2	1.9	2.3	1.2	2.3	2.3	.0	1.5
Ford	4.0	1.7	2.2	2.5	3.9	7.0	6.9	44.9	5.6	3.6	2.3	4.1	3.2	3.4	3.1
Rov	3.0	.7	.7	.6	1.8	.3	1.8	1.9	31.9	2.4	1.2	1.2	1.3	.0	.4
vw	1.0	1.8	1.7	3.6	3.7	5.5	7.6	4.2	3.6	45.9	3.5	5.3	6.8	10.3	3.8
Alfa	.0	.7	.5	.5	1.4	.6	1.8	.9	1.4	1.0	56.1	2.3	1.6	3.4	1.2
Volv	.0	.4	.3	.2	.4	.6	1.2	.3	.8	1.1	.6	45.6	1.0	6.9	1.2
BMW	.0	.7	.6	.8	.7	.6	1.2	.9	1.7	1.7	2.3	2.3	52.6	10.3	7.3
Saab	.0	.0	.0	.0	.1	.0	.0	.0	.0	.1	.6	.6	.6	34.5	.8
Merc	.0	.1	.1	.1	.2	.2	.3	.3	.8	.3	1.2	1.8	5.5	3.4	52.3

Upper case identifiers (COLUMNS): NEW *purchased car.*
Lower case identifiers (rows): Previously *owned car.*

Table 2

Reordering of Rows And Columns According to The First Axis of Correspondence Analysis of "Switchers Matrix" (Diagonal Elements Put to Zero in The Initial Matrix)

2.A Row Profiles (The sum of the elements of each row equals 100)

	REN	LADA	VOLV	MER	VW	BMW	CIT	ALFA	ROV	SAAB	GM	FORD	SEAT	FIAT	PEU
Peu	37.4	1.7	1.0	1.2	10.2	1.2	14.0	.7	2.4	.1	8.6	10.0	3.1	8.4	.0
Seat	28.7	.0	.0	.0	12.9	.0	12.9	2.0	3.0	.0	10.9	9.9	.0	9.9	9.9
Merc	16.9	.0	5.1	.0	8.5	28.8	5.1	3.4	5.1	1.7	3.4	8.5	1.7	5.1	6.8
Volv	19.8	.0	.0	3.5	20.9	3.5	12.8	1.2	3.5	2.3	7.0	4.7	4.7	5.8	10.5
Cit	32.8	1.4	.7	1.0	6.2	.9	.0	.4	2.0	.1	4.6	6.7	1.9	9.3	31.9
Rov	21.2	2.8	.9	.5	18.9	1.9	8.5	.9	.0	.0	1.4	14.2	2.8	11.8	14.2
Ford	25.1	1.4	1.2	1.4	10.3	1.7	7.7	.7	3.4	.2	13.1	.0	4.0	9.1	20.8
Fiat	24.1	2.0	.8	.5	13.4	.8	11.3	1.8	4.4	.2	8.0	8.2	2.8	.0	21.9
GM	24.8	2.2	1.2	1.2	12.1	2.2	6.2	.6	2.2	.0	.0	15.5	3.7	7.1	21.1
Alfa	21.6	.O	2.5	1.9	10.5	3.1	11.7	.0	3.1	.6	4.3	8.6	3.7	11.7	16.7
BMW	20.3	.0	2.0	9.6	14.7	.0	10.2	2.0	3.0	1.5	3.0	7.1	2.0	4.6	19.8
Saab	.0	.0	11.1	22.2	22.2	22.2	.0	11.1	.0	.0	.0	.0	.0	11.1	.0
VW	19.0	.3	1.5	1.7	.0	3.5	8.3	1.0	2.1	.5	9.7	10.7	4.1	8.4	29.2
Lada	15.5	.0	.0	.0	6.8	.0	11.7	.0	3.9	.0	11.7	10.7	2.9	12.6	24.3
Ren	.0	1.4	.9	1.2	11.1	1.3	12.6	.8	2.8	.1	9.1	11.0	3.8	11.9	32.0

2.B Column Profiles (The sum of the elements of each column equals 100)

	REN	LADA	VOLV	MER	VW	BMW	CIT	ALFA	ROV	SAAB	GM	FORD	SEAT	FIAT	PEU
Peu	38.6	26.2	21.5	18.5	21.8	15.6	30.1	17.1	19.2	10.5	23.3	22.8	20.3	19.4	.0
Seat	1.5	.0	.0	.0	1.4	.0	1.4	2.6	1.2	.0	1.5	1.2	.0	1.2	.5
Merc	.5	.0	3.2	.0	.5	11.6	.3	2.6	1.2	5.3	.3	.6	.3	.4	.2
Volv	.9	.0	.0	2.4	2.0	2.0	1.2	1.3	1.2	10.5	.8	.5	1.4	.6	.4
Cit	25.3	16.2	10.8	12.1	9.9	8.8	.0	7.9	11.8	5.3	9.3	11.5	9.5	16.1	23.2
Rov	2.4	4.6	2.2	.8	4.4	2.7	2.0	2.6	.0	.0	.4	3.5	2.0	3.0	1.5
Ford	7.7	6.2	7.5	6.5	6.6	6.8	5.0	5.3	8.2	5.3	10.6	.0	7.8	6.3	6.0
Fiat	7.8	9.2	5.4	2.4	9.0	3.4	7.6	14.5	11.0	5.3	6.8	5.9	5.8	.0	6.7
GM	4.2	5.4	4.3	3.2	4.3	4.8	2.2	2.6	2.9	.0	.0	5.9	4.1	2.7	3.4
Alfa	1.9	.0	4.3	2.4	1.9	3.4	2.1	.0	2.0	5.3	1.0	1.6	2.0	2.2	1.3
BMW	2.1	.0	4.3	15.3	3.2	.0	2.2	5.3	2.4	15.8	.8	1.6	1.4	1.1	1.9
Saab	.0	.0	1.1	1.6	.2	1.4	.0	1.3	.0	.0	.0	.0	.0	.1	.0
VW	6.1	1.5	9.7	8.1	.0	14.3	5.5	7.9	5.3	15.8	8.2	7.6	8.5	6.0	8.8
Lada	.8	.0	.0	.0	.8	.0	1.3	.0	1.6	.0	1.7	1.3	1.0	1.5	1.2
Ren	.0	30.8	25.8	26.6	34.0	25.2	39.0	28.9	31.8	21.1	35.3	36.1	35.9	39.5	44.7

Upper case identifiers (COLUMNS): NEW *purchased car.*
Lower case identifiers (rows): Previously *owned car.*