

## Initial Data Analysis

**Christopher Chatfield**

### Preliminary Comments

- a) Most statistics textbooks cover only one aspect of contingency tables; namely testing rows and columns for independence. This is clearly unnecessary here as the diagonal terms (the repeat purchases) are so large. Thus most textbooks don't help at all!
- b) The given row and column "totals" are replaced with the actual row and column totals.
- c) First question: Is the percentage of repeat purchases the same for all makes? Answer: NO. In France 89 the percentage of repeat purchases varies from 70% for Mercedes down to 26% for Seat.
- d) Second question: Are off-diagonal elements of the table in some sense independent? e.g. Is the percentage of people who previously bought make Y, who switch to make Z, proportional to the numbers buying make Z (or the numbers switching to make Z or something similar)? Answer: NO. For example 11.7% of Renaults switch to Peugeot, but only 2.1% of Mercedes switch to Peugeot.
- e) Thus there seems little alternative to giving some version of the whole table.

### Prior Information

I have no quantitative knowledge about car brand-switching. I expect repeat purchasing to be higher for quality makes and for makes with high brand share. I also expect more brand-switching between makes from the same country and of comparable price/quality. (I could be wrong).

### The Analysis

The table for France 89 is presented by giving the percentage of previous buyers of a make who buy each of the other makes. (i.e. the ratio of each frequency to the row total). Question: What ordering of rows and columns should be used? The given one is alphabetical and hardly likely to be the most helpful. I have chosen to order rows (and the equivalent columns) by the decreasing order of the percentage of repeat buyers.

### The Table

This gives a clear descriptive summary of the data. It is a good idea to spend a minute or two looking at the table.

Note (i) the clear self-explanatory title; (ii) percentages have been given to one decimal place (which is plenty!); (iii) small gaps have been inserted between every 5th column and every 5th row to aid clarity; (iv) brand shares are given; (v) Saab has been removed as the frequencies are so small.

### The Results

The brand *shares* vary enormously from 31% (Renault) down to 0.8% (Volvo, Alfa) or even less for the excluded Saab. They can be read from the table.

The *repeat purchase percentages* vary from 70% down to 26%. Mercedes is "top" despite its small brand share. The 4 makes with the largest brand shares have the next 4 highest repeat purchase values. French makes do well as we would expect in France. It is surprising (to me) that

Lada's repeat purchase proportion is larger than that of Alfa-Romeo and only one place below BMW. The full list of repeat purchase percentages and brand shares can be found from the table.

Within each column the *off-diagonal percentages* are surprisingly (?) constant. Thus in the Mercedes column the switch percentages only vary between 0.0 and 1.9 apart from the 5.3 value for BMW. Thus switching is approximately independent apart from a few exceptions which can be spotted and listed. Thus BMW → Mercedes and vice-versa and Seat → Renault are "high" but these three effects are all based on fairly small frequencies. Mercedes → Renault, Peugeot, VW and Citroen are all "low", but remember that the switch proportions at the bottom of the table are bound to be somewhat higher as the repeat purchases are smaller there. By taking percentages of switchers we could probably make off-diagonal elements even more constant in each column.

### Comparison with Other Data

I have only had time to look at France 86 and UK 89 data to get a feel of similarities/differences between years and countries.

*France 86.* Repeat purchase patterns look very similar. For example Alfa, Fiat, Lada, Rover and Seat were all still below 50%. Switch patterns also look similar. The BMW → Mercedes link held in 86 as well.

*UK 89.* There is less variability in repeat purchase probabilities than in France. All are in the 50-60% range except Renault (46%; much lower than in France) and Volvo (61%; higher than France), Ford (64%; higher than France), Mercedes (65%), and Nissan (68%). Switch patterns also exhibit approximate independence in general but there are rather more exceptions than in France. The BMW → Mercedes link is present in UK as well as France, and there is higher than normal switching between GM → Ford, Rover → Ford (an English link), Toyota → Honda (a Japanese link), and Saab → Volvo (a Swedish link).

**Table 1**

FRANCE 1989 Percentages of buyers of previous make who bought each of 14 makes, together with totals and new brand shares. Makes are arranged in order of the percentage repurchasing same make.

PREVIOUS MAKE	NEWLY PURCHASED MAKE														Total Sample
	Merc	Ren	Peu	VW	Cit	Ford	GM	Volv	Fiat	BMW	Lada	Alfa	Rov	Seat	
Mercedes	70.1	5.2	2.1	2.6	1.5	2.6	1.0	1.5	1.5	8.8	0.0	1.0	1.5	0.5	194
Renault	0.4	63.5	11.7	4.0	4.6	4.0	3.3	0.3	4.4	0.5	0.5	0.3	1.0	1.4	7656
Peugeot	0.5	14.9	60.1	4.1	5.6	4.0	3.4	0.4	3.4	0.5	0.7	0.3	1.0	1.2	4875
VW/Audi	0.7	8.4	12.9	56.1	3.6	4.7	4.3	0.7	3.7	1.5	0.1	0.4	0.9	1.8	1375
Citroen	0.5	14.6	14.2	2.8	55.5	3.0	2.1	0.3	4.2	0.4	0.6	0.2	0.9	0.9	3265
Ford	0.6	11.4	9.5	4.7	3.5	54.5	6.0	0.5	4.2	0.8	0.6	0.3	1.6	1.8	1277
GM	0.6	11.7	9.9	5.7	2.9	7.3	52.8	0.6	3.4	1.0	1.0	0.3	1.0	1.8	685
Volvo	1.9	10.5	5.6	11.1	6.8	2.5	3.7	48.1	3.1	1.9	0.0	0.6	1.9	2.5	162
Fiat	0.3	13.0	11.8	7.2	6.1	4.4	4.3	0.4	46.2	0.4	1.1	1.0	2.4	1.5	1138
BMW	5.3	11.2	10.9	8.1	5.6	3.0	1.7	1.1	2.5	45.7	0.0	1.1	1.7	1.1	357
Lada	0.0	9.4	14.6	4.1	7.0	6.4	7.0	0.0	7.6	0.0	39.8	0.0	2.3	1.8	171
Alfa Romeo	1.2	13.6	10.5	6.6	7.4	5.4	2.7	1.6	7.4	1.9	0.0	37.6	1.9	2.3	258
Rover	0.3	13.8	9.2	12.2	5.5	9.2	0.9	0.6	7.6	1.2	1.8	0.6	35.2	1.8	327
Seat	0.0	21.2	7.3	9.5	9.5	7.3	8.0	0.0	7.3	0.0	0.0	1.5	2.2	26.3	137
Total Sample	258	6747	4932	1681	2717	1550	1082	170	1371	308	198	172	360	331	21877
Brand Share %	1.2	30.8	22.5	7.7	12.4	7.1	4.9	0.8	6.3	1.4	0.9	0.8	1.6	1.5	100.0