Understanding household survey nonresponse through geo-demographic coding schemes

King, John

Understanding Household Survey Nonresponse through Geo-demographic Coding Schemes[1]

JOHN KING

Abstract: A geo-demographic coding system was used to analyse response to a large household survey. This provided a contemporaneous indicator of response and of nonresponse to the survey using terms external to the survey. Geo-demographic coding information is used to describe the non-respondents to the survey. Use of this information for reducing some fieldwork problems is also indicated.

Keywords: household sample surveys; survey nonresponse; geo-demographic coding systems; non-respondent characteristics; Mosaic.

1   Background

Geo-demographic coding systems have been used extensively for marketing and planning. In partnership with Experian, owners of the Mosaic geo-demographic coding system, Mosaic codes were attached to the eligible sample drawn for the British Family Expenditure Survey (FES) for 1995-96. Sample response and nonresponse was examined using these codes. Usually response to surveys is analysed either by sample design or fieldwork characteristics; or by socio-economic descriptors. The geo-demographic codes have given a description of survey nonresponse using an alternative set of descriptors. A simple, but basically multi-factor, picture of respondents and non-respondents can be obtained using geo-demographic coding schemes.

A previous paper (King 1996) indicated ways in which these schemes could be used to describe survey respondents. This paper extends those analyses to the characteristics of the eligible drawn sample. This leads to a description, in geo-demographic terms, of the non-respondents as well as of the respondents.

2   Geo-demographic coding schemes

Geo-demographic coding schemes have been developed over the last couple of decades to help with analyses in several subject areas; particularly in market research, and especially targeted marketing, and in urban social geography and planning. There are several major schemes available in Great Britain based on the 1991 Census of Population and Housing. A brief description of these, and the principles upon which they are built, is given by Openshaw (1993).

The coding of households by geo-demographic coding schemes is not exact, in the sense that the characteristics of the individual household are not used. First, the coding scheme uses information from different geographic levels. Some information may be specific to the unit postcode, some information may be averaged over several adjacent unit postcodes and other information may refer to a wider geographic area. Second, all households with a particular postcode will be given the same code; no allowance is made for the characteristics of the individual household. By using the geo-demographic codes in this study, it is assumed that the codes do adequately describe the responding households.

The system used for this study, and the previous one, is the Mosaic system: a product of Experian which is used here with their permission. The Mosaic system is based on 86 variables and is described in more detail in the earlier paper and in The Multimedia Guide to Mosaic (Experian 1997). Essentially, about half of the input variables are Census variables (or census-based) which are available at the level of the census Enumeration District; other variables, including some financial information and data on accessibility to shopping centres, are available at the level of the postcode.

The Mosaic classification system has 52 household types. These are grouped into 12 lifestyle groupings as shown in Annex A. Households are allocated a Mosaic type using only the postcode of the household.

3   The survey and attaching the codes

The Family Expenditure Survey (FES) is a continuous survey of private households conducted in the United Kingdom by the Office for National Statistics (ONS). Detailed information about the expenditure of all those aged 16 and over is obtained through interviews and a diary kept for a period of 2 weeks. Information about income of individuals in the household is also obtained through interviews, along with information about employment, receipt of state security benefits and ownership of consumer durable goods. Only the sample for Great Britain has been used for this study.

In Great Britain the sampling frame is the small-user Postcode address file (PAF) maintained by the Post Office. The design is a multi-stage stratified systematic sample with clusters. The sample is stratified by Standard Statistical Region (SSR) and by three further variables: metropolitan/non-metropolitan areas (the latter are further split into 3 strata of low, medium and high population density); the proportion of owner-occupiers and the proportion of privately renting households.

The Social Survey Division (SSD) of ONS draws the FES sample and undertakes the fieldwork in Great Britain. Information is collected under a promise of confidentiality. To ensure this, any names and addresses are removed before processing, and records are further anonymised before being released as micro-data. Attaching a geo-demographic code is not thought to compromise this undertaking.

The Mosaic code for a household is determined solely by its postcode. SSD used a directory of Mosaic codes provided by Experian to Mosaic-code the drawn FES sample. This was subsequently modified to a list of the eligible sample by excluding ineligible addresses and adding those additional households sampled at addresses where multiple households were found. This list also showed the outcome of the fieldwork: response, refusal or non-contact.

4   The FES sample by geo-demographic codes

Table 1 shows the eligible sample by Mosaic main lifestyle groupings for 1995-96 and also the composition of the household population in Mosaic terms. The distribution of FES households across these groups is compared with the distribution of all households. The Index in the final column of Table 1 is the ratio of FES households to all households: an Index value of 100 means that there was the same proportional representation in the FES as in the population.

Differences between the two distributions are attributable to several factors. One factor is the slightly varying definition of a household: the FES uses the pre-1981 census household definition, based on sharing common housekeeping and meals; while the Mosaic system was built using the 1981, and subsequent, household definition based only on the criterion of shared accommodation. Estimates of the distribution of households by Mosaic type are updated using Electoral Roll and PAF data. The differences created by the change in Census definitions are thought to be small (Todd and Griffiths 1986). Other differences are due to sampling variability and to changes over time from the 1991 Census basis of some of the variables underlying the Mosaic classification system.

Table 1: Family expenditure survey 1995-96 - Great Britain
The eligible sample by Mosaic codes

Lifestyle grouping		Number of eligible FES households	Percentage of households in		Index
			FES	**Mosaic**	
L1	High Income Families	1171	11.7	9.9	118
L2	Suburban Semis	1162	11.6	11.0	105
L3	Blue Collar Owners	1297	12.9	13.0	99
L4	Low Rise Council	1482	14.7	14.4	102
L5	Council Flats	695	6.9	6.8	102
L6	Victorian Low Status	822	8.2	9.4	87
L7	Town Houses & Flats	974	9.7	9.4	103
L8	Stylish Singles	489	4.9	5.2	94
L9	Independent Elders	720	7.2	7.4	97
L10	Mortgaged Families	550	5.5	6.2	88
L11	Country Dwellers	654	6.5	7.0	93
L12	Institutional Areas	29	0.3	0.3	96
	not Mosaic coded	109			
	Total	10154			

5   Response by geo-demographic codes - the results

Table 2 shows the number of households in the FES in 1995-96 in Great Britain in each of the Mosaic main lifestyle groupings by the outcome of the fieldwork: full response, refusal and non-contact. Annex B shows similar information for each of the 52 Mosaic types.

Table 2 shows that the response rate for the FES in GB in 1995-96 was 65 per cent. Nearly all the nonresponse was due to refusals rather than non-contact. Refusals were 32 per cent and non-contacts were 3 per cent.

**Table 2:   Family expenditure survey 1995-96 - Great Britain
              Analysis of the results of the fieldwork by Mosaic codes**

Lifestyle grouping		Number of eligible FES households			
		responding	refusing	non-contacted	total
L1	High Income Families	776	376	19	1171
L2	Suburban Semis	781	364	17	1162
L3	Blue Collar Owners	843	422	32	1297
L4	Low Rise Council	991	457	34	1482
L5	Council Flats	411	245	39	695
L6	Victorian Low Status	494	280	48	822
L7	Town Houses & Flats	630	322	22	974
L8	Stylish Singles	282	164	43	489
L9	Independent Elders	465	237	18	720
L10	Mortgaged Families	386	146	18	550
L11	Country Dwellers	458	190	6	654
L12	Institutional Areas	24	4	1	29
not Mosaic coded		70	38	1	109
	Total	6611	3245	298	10154

Table 3 shows the distributions across the Mosaic groupings for the respondents and non-respondents. The distribution of responding households differs from the eligible sample shown in Table 1 mainly in that there were significantly fewer responding households in groups L5 Council Flats, L6 Victorian Low Status and L8 Stylish Singles. This is a similar pattern to that found in the analysis of respondents to the 1994-95 FES (King 1996).

Table 3 also shows the characteristics of the non-responding households. This is direct description of these households. There are proportionately more non-responding households in groups L5, L6 and L8, corresponding to the under-representation of these groups in the responding households. There are also proportionately fewer households in the groups L2 Suburban Semis, L4 Low Rise Council, L10 Mortgaged Families and L11 Country Dwellers. Only four groups have a representation similar to that in the eligible sample.

The type of nonresponse - either refusal or non-contact - is known for each household. Table 3 also shows the characteristics of household by reason for nonresponse. Non-contact is low on the FES and the number of households not contacted is small. This is because of the efforts of the interviewers and also because the design of the sample and the field work permits several calls. But although the number of these households is relatively low, the characteristics of these are very different from those of the refusing households and of the eligible sample. In particular, again, households in groups L5, L6 and L8 are proportionately over-represented. These categories are known to be problematic: the access to many council flats is difficult because of the security arrangements; and stylish singles are often difficult to find at home.

Table 3: Family expenditure survey 1995-96 - Great Britain
Response and nonresponse by Mosaic codes

		Percentage of eligible FES households			
Lifestyle grouping		responding	nonresponse (total)	refused	no contact
L1	High Income Families	11.9	11.3	11.7	6.4
L2	Suburban Semis	11.9	10.9	11.4	5.7
L3	Blue collar Owners	12.9	13.0	13.2	10.8
L4	Low Rise Council	15.2	14.0	14.3	11.4
L5	Council Flats	6.3	8.1	7.6	13.1
L6	Victorian Low Status	7.6	9.4	8.7	16.2
L7	Town Houses & Flats	9.6	9.8	10.0	7.4
L8	Stylish Singles	4.3	5.9	5.1	14.5
L9	Independent Elders	7.1	7.3	7.4	6.1
L10	Mortgaged Families	5.9	4.7	4.6	6.1
L11	Country Dwellers	7.0	5.6	5.9	2.0
L12	Institutional Areas	0.4	0.1	0.1	0.3
	Total Mosaic coded	6541	3504	3207	297

Table 4 shows another way of looking at the data. For each Lifestyle grouping the row shows the proportions of households responding, refusing and not being contacted. Thus, L10 and L11 are groups with higher response. The groups most difficult to contact are L5, L6 and L8. These have only a slightly higher refusal rate than other household types: the main difficulty is in contacting them. Contrary to general beliefs, L1 High Income Families are not poor participants in the FES. Their response rate is about average, and higher than that of several other groups.

**Table 4:   Family expenditure survey 1995-96 - Great Britain
             Response and nonresponse patterns by Mosaic codes**

Lifestyle grouping		Percentage of eligible FES households			
		responding	refusing	no contact	Total (number)
L1	High Income Families	66	32	2	1171
L2	Suburban Semis	67	31	1	1162
L3	Blue collar Owners	65	33	2	1297
L4	Low Rise Council	67	31	2	1482
L5	Council Flats	59	35	6	695
L6	Victorian Low Status	60	34	6	822
L7	Town Houses & Flats	65	33	2	974
L8	Stylish Singles	58	34	9	489
L9	Independent Elders	65	33	3	720
L10	Mortgaged Families	70	27	3	550
L11	Country Dwellers	70	29	1	654
L12	Institutional Areas	83	14	3	29
	Total Mosaic coded	65	32	3	10045

6   Comparisons with other analyses

Foster (1996) analysed FES respondents and non-respondents using certain census variables, and described the factors most closely affecting nonresponse in terms of these variables. This was based on an exercise matching the FES sample with individual household records from the 1991 Census of Population and Housing. The number of adults in the household, especially where there were no children under the age of 16, was the variable most closely associated with response, probably because of the strict response rules for the FES. Other characteristics linked with higher nonresponse were: if the household was in London; if the household had a head with no post-school qualifications, was born outside the UK or was self-employed. Also, nonresponse was associated with households with no dependent children or with a head of household aged 35 or more.

Because the categories used for Foster's analysis were different from those used in this study it is not possible to make direct comparisons between the conclusions of the studies. However, there are some indications of these results. The high representation of L5 and

L6 groups in the non-responding households is reflected in the finding about London and other metropolitan areas. Similarly, the higher representation of L8 is part of the description of households with no dependent children.

Analysis by geo-demographic coding schemes will not replace the need for Census comparisons to understand the true picture of response and nonresponse. But such studies are infrequent and costly. By comparison, geo-demographic coding analyses are cheap and very fast. They could be used to give a cheap means of monitoring trends in nonresponse. Monitoring trends would provide checks on the continued appropriateness of the assumptions made when using Census-based studies several years after the Census date.

A more recent study of nonresponse to the FES has been undertaken by Hansbro and Foster (1997). This study reports on an analysis of a nonresponse questionnaire (NRQ). The NRQ is used by SSD to obtain as much information as possible about non-responding households on certain key descriptive variables. Again, the analysis is in terms of standard, Census-like, socio-economic descriptors. Much of the picture emerging from this study is similar to the Census comparison and to the implicit results of the geo-demographic code analysis.

7   Value and uses of geo-demographic analyses

A particular advantage of geo-demographic coding is that information is available quickly and relatively cheaply. There is a built-in picture of nonresponse as the data flows in from the field. It is thus possible to monitor levels and trends in response immediately. Month by month tracking of trends would allow problems to be seen and even anticipated.

Geo-demographic coding information for each sampled household is, of course, available as soon as the sample is drawn. This means it can be used by fieldwork organisers to anticipate the likely difficulty of a workload. Allocation of interviewers can be targeted; matching interviewers with particular skills to particular workloads - the nature of which will be better known.

Interviewer performance and variability can also be understood in the context of the actual workload and the difficulty of it. Geo-demographic codes provide a more objective measure of the difficulties and problems that the interviewers may have faced.

8   The future

We hope to look further at the stability of these analyses over time. We hope to look at the bias resulting from nonresponse in both expenditure and income patterns. We also hope to compare in more detail these results with the traditional studies mentioned above.

References

Experian (1997). The Multimedia Guide to MOSAIC. Nottingham: Experian

Foster, K. (1996). A comparison of the census characteristics of respondents and non-respondents to the 1991 Family Expenditure Survey. London: Office for National Statistics; Survey Methodology Bulletin, no 38, January 1996

Hansbro, J. and Foster, K. (1997). Characteristics of nonresponse on the 1995-96 Family Expenditure Survey. London: Office for National Statistics; Social Survey Division

King, J. (1996). The use of geo-demographic coding schemes for understanding household survey nonresponse. Seventh International Workshop on Household Survey Nonresponse. Rome, October 1996

Openshaw, S. (1993). Special Classifications. Chapter 7 in: B. Leventhal, C. Moy and J. Griffin (eds.). An Introductory Guide to the 1991 Census. Henley-on-Thames: NTC Publication Ltd in association with the Market Research Society

Todd, J. and Griffiths, D. (1986). Changing the definition of a household. London: Office for National Statistics; Social Survey Division

Annex A
Structure of the Mosaic coding system

L1 High Income Families
    M1   Clever Capitalists
    M2   Rising Materialists
    M3   Corporate Careerists
    M4   Ageing Professionals
    M5   Small Time Business
L2 Suburban Semis
    M6   Green Belt Expansion
    M7   Suburban Mock Tudor
    M8   Pebble Dash Subtopia
L3 Blue Collar Owners
    M9   Affluent Blue Collar
    M10   30s Industrial Spec
    M11   Lo-Rise Right to Buy
    M12   Smokestack shiftwork
L4 Low Rise Council
    M13   Coop club and Colliery
    M14   Better off Council
    M15   Low Rise Pensioners
    M16   Low Rise Subsistence
    M17   Problem Families
L5 Council Flats
    M18   Families in the Sky
    M19   Graffitied Ghettos
    M20   Small Town Industry
    M21   Mid Rise Overspill
    M22   Flats For The Aged
    M23   Inner City Towers
L6 Victorian Low Status
    M24   Bohemiam Melting Pot
    M25   Victorian Tenements
    M26   Rootless Renters
    M27   Sweatshop Sharers
    M28   Depopulated Terraces
    M29   Rejuvenated Terraces

L7 Town Houses & Flats
    M30   Bijou Homemakers
    M31   Market Town Mixture
    M32   Town Centre Singles
L8 Stylish Singles
    M33   Bedsits and Shop Flats
    M34   Studio Singles
    M35   College and Communal
    M36   Chattering Classes
L9 Independent Elders
    M37   Solo Pensioners
    M38   High Spending Greys
    M39   Aged Owner Occupiers
    M40   Elderly in Own Flats
L10 Mortgaged Families
    M41   Brand New Areas
    M42   Pre Nuptial Owners
    M43   Nestmaking Families
    M44   Maturing Mortgagees
L11 Country Dwellers
    M45   Gentrified Villages
    M46   Rural Retirement Mix
    M47   Lowlands Agribusiness
    M48   Rural Disadvantage
    M49   Tied / Tenant Farmers
    M50   Upland and Small Farm
L12 Institutional Areas
    M51   Military Bases
    M52   Non Private Housing

Mosaic codes are used by permission of Experian.

Annex B
Family expenditure survey 1995-96 - Great Britain
Response type by Mosaic codes

	Description	respondents	non-respondents	refusals	no contact	Total (number)
M1	Clever Capitalists	1.4	1.8	1.7	2.0	153
M2	Rising Materialists	2.8	1.9	1.9	2.0	247
M3	Corporate Careerists	2.9	2.9	3.1	0.7	291
M4	Ageing Professionals	1.6	1.7	1.8	0.7	165
M5	Small Time Business	3.2	3.0	3.2	1.0	315
M6	Green Belt Expansion	3.9	2.6	2.7	1.7	347
M7	Suburban Mock Tudor	3.1	3.0	3.1	2.4	310
M8	Pebble Dash Subtopia	4.9	5.2	5.6	1.7	505
M9	Affluent Blue Collar	2.9	3.5	3.6	2.4	312
M10	30s Industrial Spec	4.0	4.0	4.1	2.4	398
M11	Lo-Rise Right to Buy	3.0	3.1	3.0	3.7	304
M12	Smokestack shiftwork	3.0	2.4	2.4	2.4	283
M13	Coop club and Colliery	3.6	3.1	3.3	1.0	341
M14	Better off Council	2.0	2.0	2.0	2.4	204
M15	Low Rise Pensioners	4.0	3.9	4.1	1.7	399
M16	Low Rise Subsistence	3.7	3.3	3.1	4.7	357
M17	Problem Families	1.8	1.7	1.7	1.7	181
M18	Families in the Sky	1.2	1.8	1.6	4.7	141
M19	Graffitied Ghettos	0.3	0.3	0.2	0.3	27
M20	Small Town Industry	1.1	1.2	1.2	1.0	115
M21	Mid Rise Overspill	0.6	0.6	0.6	0.3	58
M22	Flats for the Aged	1.5	1.6	1.7	1.0	152
M23	Inner City Towers	1.7	2.6	2.3	5.7	202
M24	Bohemiam Melting Pot	1.8	3.0	2.6	7.1	221
M25	Victorian Tenements	0.0	0.2	0.1	1.0	7
M26	Rootless Renters	0.8	0.9	0.9	1.3	87
M27	Sweatshop Sharers	0.9	0.9	0.9	1.7	92
M28	Depopulated Terraces	0.6	0.8	0.9	0.3	70
M29	Rejuvenated Terraces	3.4	3.5	3.3	4.7	345
M30	Bijou Homemakers	3.4	3.9	4.0	2.4	358
M31	Market Town Mixture	4.2	3.5	3.7	1.7	397

Percentage of eligible FES households

	Description	Percentage of eligible FES households				
		respondents	non-respondents	refusals	no contact	Total (number)
M32	Town Centre Singles	2.0	2.4	2.3	3.4	219
M33	Bedsits and Shop Flats	0.7	1.1	0.9	2.7	82
M34	Studio Singles	1.3	2.4	2.0	7.4	172
M35	College and Communal	0.5	0.4	0.3	1.3	45
M36	Chattering Classes	1.8	2.0	1.9	3.0	190
M37	Solo Pensioners	2.0	2.1	2.1	2.4	206
M38	High Spending Greys	0.9	0.9	1.0	0.3	93
M39	Aged Owner Occupiers	2.9	2.5	2.5	1.7	278
M40	Elderly in Own Flats	1.3	1.7	1.7	1.7	143
M41	Brand New Areas	0.2	0.2	0.2	0.3	20
M42	Pre Nuptial Owners	1.4	0.6	0.5	1.3	113
M43	Nestmaking Families	1.8	1.6	1.5	2.0	174
M44	Maturing Mortgagees	2.5	2.3	2.3	2.4	243
M45	Gentrified Villages	1.5	1.5	1.6	0.0	150
M46	Rural Retirement Mix	0.5	0.3	0.2	0.3	43
M47	Lowlands Agribusiness	1.6	1.7	1.7	1.3	164
M48	Rural Disadvantage	1.2	0.9	0.9	0.3	110
M49	Tied / Tenant Farmers	0.8	0.7	0.7	0.0	73
M50	Upland and Small Farm	1.4	0.7	0.7	0.0	114
M51	Military Bases	0.4	0.1	0.1	0.3	29
M52	Non Private Housing	0.0	0.0	0.0	0.0	0
	Total Mosaic coded	6541	3504	3207	297	10045
	Households not coded	70	39	38	1	109
	Total all Households in FES	6611	3543	3245	298	10154