Statistical Briefing Note No. 2

Bias in the Northern Ireland Omnibus June 2012 sort card module?

David Gordon, Shailen Nandy, Demi Patsios, University of Bristol
December 2012

Background

Two different approaches were used to gather information on necessities in Northern Ireland Omnibus Survey in June 2012: (1) a Personal Digital Assistant 'PDA' method which involved handing the PDA to the respondent so that they could answer the necessities questions, this was the method used in the 2002/03 *Bare Necessities* Survey in NI, and (2) the original 'sort card' method first used by Mack and Lansley in the 1983 *Poor Britain Survey*.

For the June 2012 survey, a systematic random sample of 2,200 addresses were selected from the Land and Property Services Agency list of private addresses. Each address was then assigned a serial number: cases with an odd serial number were selected for the 'sort card' version of the Necessities module, while those cases with an even serial number were routed to the 'PDA' version of the Necessities module. According to NISRA, this methodology ensured that both samples were representative.[1]

Unfortunately, a programming error resulted in a hard check, which has been disabled during the survey development phase, not being applied to the sort card sample, which meant that the interviewers were not prompted if they did not enter a valid response for all the sort cards. This resulted in a higher than expected number of missing values for the individual activities and items covered in the Necessities module. These were in addition to the 16 respondents who refused to answer any of the Necessities module questions. A number of sort cards were unaccounted for as a result of recording and/or keying errors by the interviewers. These errors make up the majority of missing items. NISRA carried out a complete manual check on the sort card sample and minor amendments to the dataset were subsequently made. NISRA states that the integrity of the data has not been unduly affected.

Aim

This report aims to test if the data from the 'sort card' sample of the 2012 NI Omnibus survey are biased as a result of interviewer recording errors? Or if these data are unbiased and can thus be used in analyses of perception of necessities in NI and comparisons with necessities results from Britain.

[1] Additional information on the Necessities module, as well as sampling and response rates can be found in: *Northern Ireland Omnibus Survey June 2012 Necessities of Life Module Codebook & Technical Summary*.

NI Omnibus Necessities Module

The NI June 2012 Omnibus survey necessities 'sort card' module asked respondents to complete the following task;

""*On these cards are a number of different items which relate to our standard of living. I would like you to indicate the living standards you feel all adults should have in Northern Ireland today by placing the cards in the appropriate box.*

Box A is for items which you think are necessary - which all adults should be able to afford and which they should not have to do without.

Box B is for items which may be desirable but are not necessary."

Respondents were handed four sets of numbered colour coded cards (one set at a time), each set of cards was shuffled to ensure that they were in a random order. Once the respondent had placed all the cards from a set into the two boxes the interviewer picked up the cards in Box A and entered each card number into their computer, they then input the numbers of cards in Box B.

Sometimes respondents may not wish to place some of the cards in either Box A or B (i.e. they cannot decide if an item is a necessity or not) they usually put these cards down by the side of the two boxes and the interviewer is meant to record these cards as 'Unallocated/ Don't Know'. A hard software check is meant to prevent the interviewer from moving to the next question until all the cards have been accounted for (i.e. all the card numbers have been entered into the computer). This check was erroneously not enabled resulting in the missing data.

A priori it seems likely that the missing data may be predominantly from 'unallocated' cards rather than from cards placed in Box A or Box B – as interviewers may have assumed that the computer would automatically fill in the 'unallocated/don't knows' once they had entered all the numbers of all the cards placed into Box A and B.

If this assumption is correct then it would be expected that the missing data will have a similar data pattern to the 'unallocated' cards which were correctly recorded by the interviewers.

Types of Missing Data

Missing data are a fact of life in survey research there is unlikely ever to have been a large social survey which did not contain some missing data. In their classic book, Little and Rubin (1987) identify three kinds of missing data;

1) Missing Completely At Random (MCAR) i.e. the probability that an item of data is missing is unrelated to its 'true' value – the value it would have had if it were not missing - or to the value of any other variable. With data that are MCAR all items/values are equally likely to be missing. The missing data are just a random sub-set of the non-missing data.

2) Missing At Random (MAR) i.e. the probability that an item of data is missing is unrelated to its 'true' value – the value it would have had if it were not missing - after controlling for another variable. With data that are MAR, missingness is correlated with other variables that if included in the analysis will provide unbiased results. Data which are MAR are *ignorable missing* once they have been controlled for.

Howell (2007) provides the following example; people who are depressed might be less inclined to report their income, and thus reported income will be related to depression. Depressed people might also have a lower income in general, and thus when we have a high rate of missing data among depressed individuals, the existing mean income might be lower than it would be without missing data. However, if, within depressed patients the probability

of reported income was unrelated to income level, then the data would be considered MAR, though not MCAR. An analysis which controlled for depression would provide unbiased income results.

3) Not Missing At Random (NMAR) i.e. the data are missing for a specific reason and estimates will be biased even after controlling for other available variables. This kind of missing data is non-ignorable.

Findings

If the missing data mainly consisted of unrecorded cards which respondents had not placed in either Box A (necessities) or Box B (Not a Necessity) then the missing data should have a similar pattern to the Don't Knows which were correctly recorded. Figure 1 shows that the Don't Knows display a clear inverse relationship when plotted against the percent of respondents who said an item was a necessity. The higher the number of respondents saying an item is a necessity the fewer the number of don't knows.

Figure 1: Percent necessities by percent Don't Knows - Adults

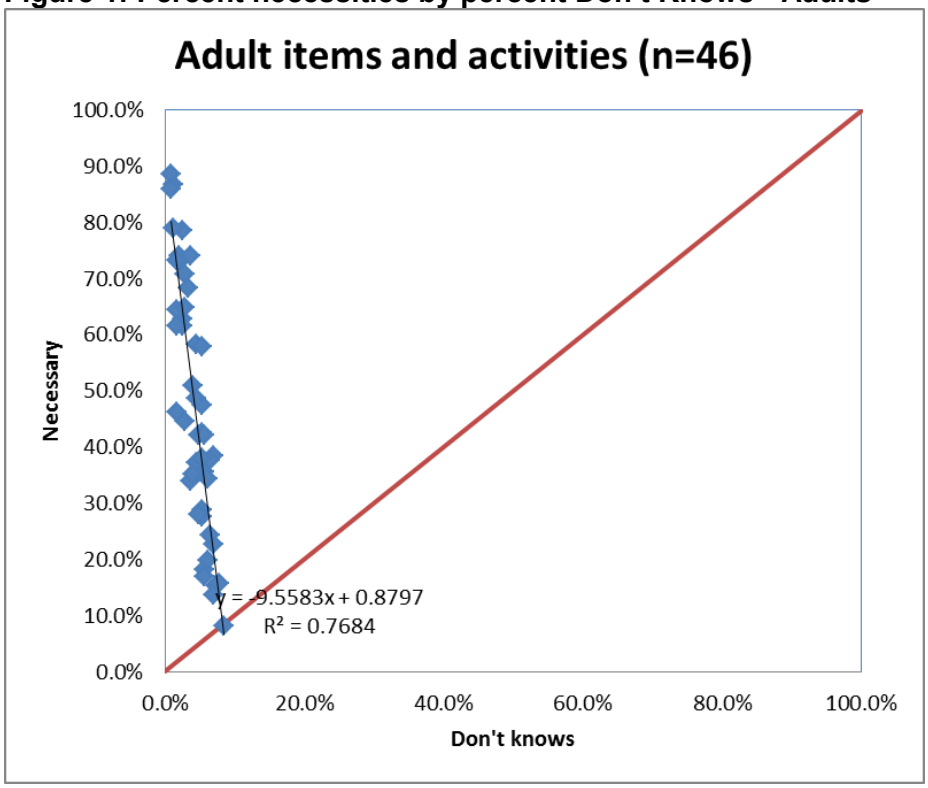


Figure 2 shows that the missing data (labelled as 'The Real 99s') display a similar inverse relationship to the Don't Know when plotted against the percent of respondents who said an item was a necessity. The higher the number of respondents saying an item is a necessity the fewer the number of missing data.

Figure 2: Percent necessities by percent missing data - Adults

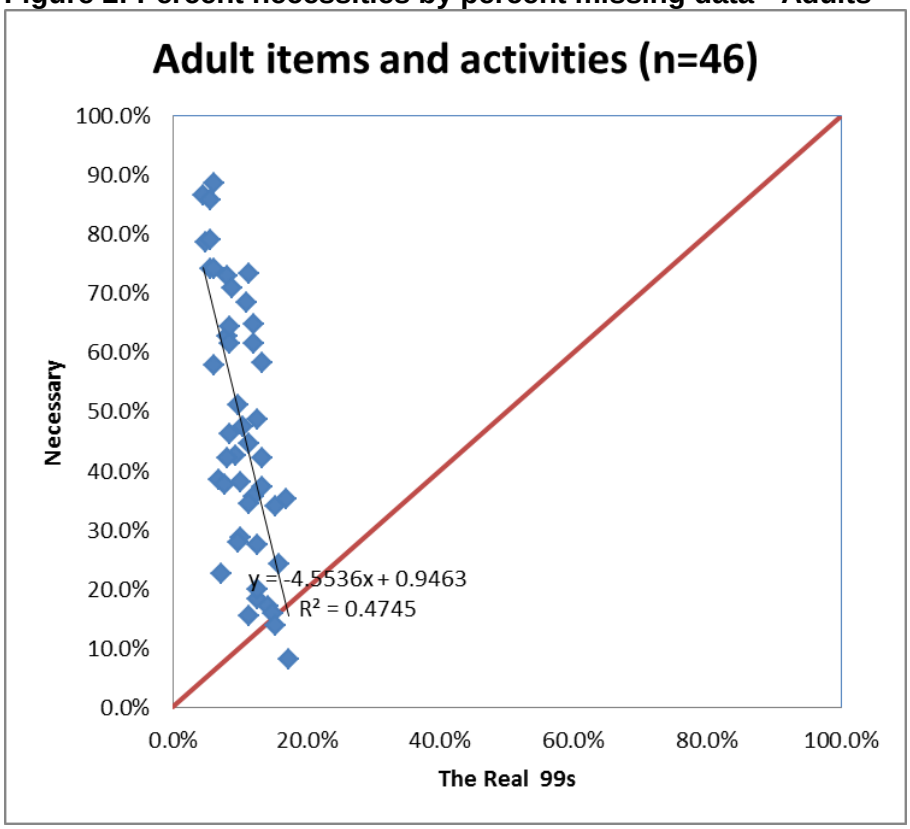


Figure 3 shows that for Children's items and activities the Don't Knows also display a clear inverse relationship when plotted against the percent of respondents who said a children's item was a necessity. The higher the number of respondents saying an children's item is a necessity the fewer the number of don't knows.

Figure 4 shows that the missing data for children's items and activities (labelled as 'The Real 99s') display a similar inverse relationship to the Don't Know when plotted against the percent of respondents who said a children's item was a necessity. The higher the number of respondents saying an item is a necessity the fewer the number of missing data.

Figure 3: Percent necessities by percent Don't Knows - Children

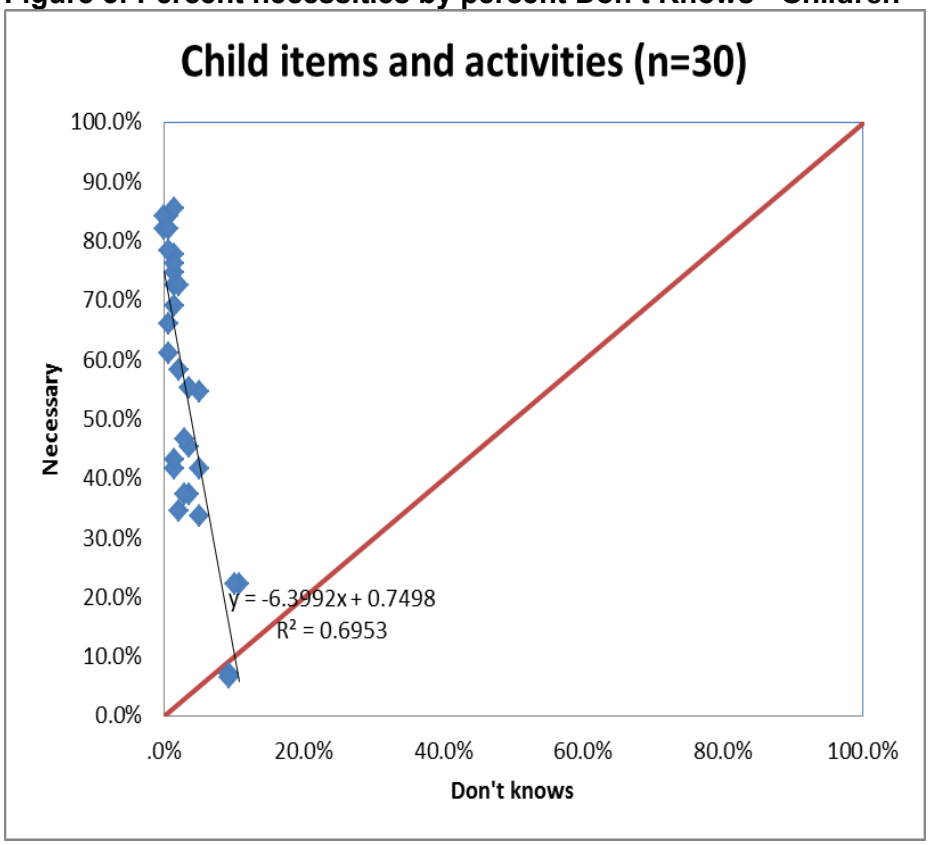
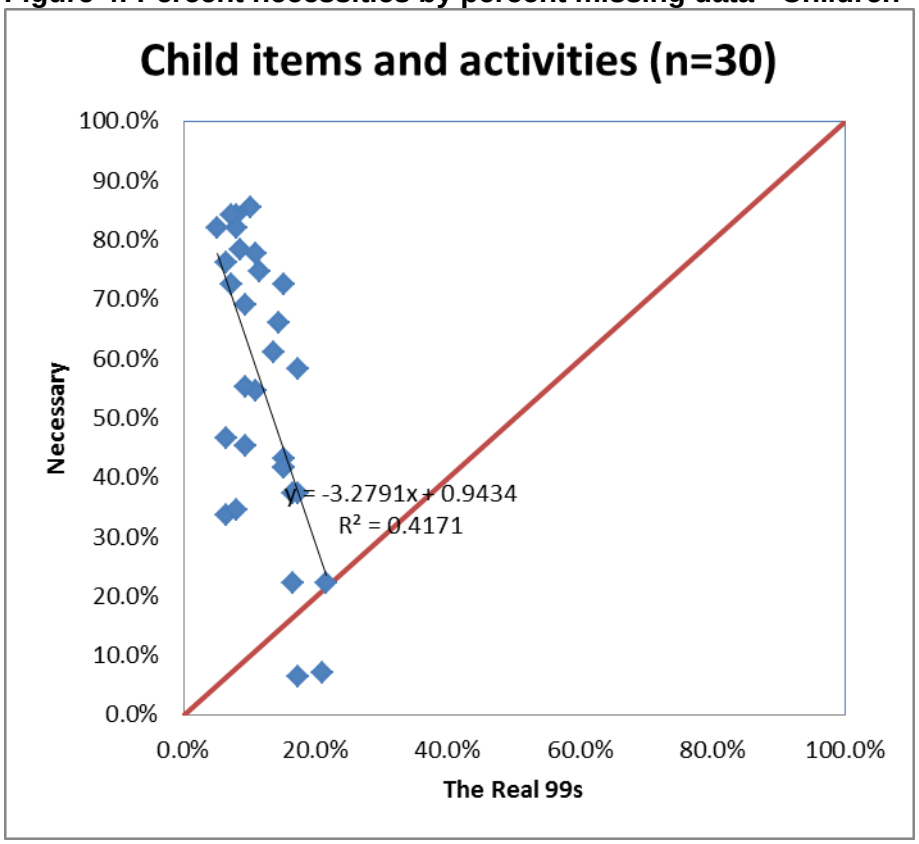


Figure 4: Percent necessities by percent missing data - Children



The results shown in Figure 1 to 4 indicate that the missing data seem likely to mainly consist of unrecorded don't knows. However it is also important to test if there are any significant

differences in the answers provided by respondents who had no missing data compared with respondents who had some missing data. If respondents without any missing data are more likely to consider an item to be a necessity than respondents who had some missing data then this may indicate that these survey data are biased as a result of the interviewer errors.

The sample was divided into two groups – Respondent with no missing data (NONE) Vs Respondents with (SOME) missing data and relative risk ratios[2] and their 95% Confidence Intervals (CIs) were calculated

Table 1: RESPONDENTS (NONE VS. SOME) MISSING (N=76 items in total)

	RR	95% CI Lo	95% CI Hi
ADULT ITEMS & ACTIVITIES (COMBINED)			
meatfish Meat, fish or vegetarian equivalent every other day (pov1_19)	0.902	0.815	0.999
heating Heating to keep home adequately warm (pov1_21)	1.037	1.005	1.071
CHILD ITEMS & ACTIVITIES (COMBINED)	NA	NA	NA

Table 1 shows that only two out of 76 relative risks are likely to be significantly different (the Confidence Intervals to not include 1.0) for 'Meatfish' (Meat, fish or vegetarian equivalent every other day) and 'heating' (Heating to keep home adequately warm). However, as these analyses involved 76 separate statistical tests it would be expected that 3 or 4 would be significant at the <5% level. Thus these results show no indication of bias in the data as a result of the interviewer errors in not correctly recording some 'necessities' item and activities cards.

Additional analyses were carried out to see if there were any systematic differences amongst demographic groups in there likelihood a valid answer (Necessary, Desirable or DK) being recorded by the interviewer Vs the answer being 'Missing'. Table 2 summarises the relative risk results for each item and activity for eight demographic groupings (Men Vs Women, Young Vs Old, Employed Vs Not Employed, High Vs Low Social Class, High Vs Low Education, Catholic Vs Protestant, Disabled Vs Not Disabled and Dependants Vs No Dependents).

In total 608 separate tests were calculated (76 items * 8 groups = 608) and it would be expected that 30 (or 31) of these tests would be 'significant' at the <5% level due to random chance. Table 2 shows that 31 of these tests were possibly significant at the <5% level, so no bias is evident due to interviewer recording errors by demographic group.

[2] A relative risk of 2.0 means twice the risk, a score of 0.5 means half the risk, a score of 3.0 is three times the risk and a score of 0.33 is a third of the risk, etc. If the 95% Confidence Intervals of a relative risk ratio span 1.0 then you cannot be confident at the 5% level that the 'true' risk is different from 50:50, i.e. the difference is unlikely to be 'significant'. Thus relative risk ratios and their 95% Confidence Intervals (CI) provide intuitive and useful estimates about whether the differences between two groups are likely to be significant (or not). For further information see, Gordon (2012) *PSE Statistical Briefing Note No 1*

Table 2: RESPONDENTS (NONE VS. SOME) MISSING BY DEMOGRAPHIC GROUP

Demographic groups	RR	95% CI Lo	95% CI Hi
(1) Male (versus Female)			
dishwash Dishwasher (pov1_25)	0.938	0.885	0.995
table A table, with chairs, at which all the family can eat (pov1_31)	0.936	0.891	0.984
worship Attending church, mosque, synagogue or other places of worship (pov2_12)	0.948	0.902	0.996
cpc Computer and internet for homework (pov3_21)	0.947	0.905	0.991
(2) 16-24 (versus 65+)	NA	NA	NA
(3) In paid employment (versus Not in paid employment)			
decorate Enough money to keep your home in a decent state of decoration (pov1_01)	0.948	0.913	0.985
nodamp Damp-free home (pov1_13)	0.970	0.943	0.997
cgames Indoor games suitable for their ages (building blocks, board games, computer games, etc.) (pov3_12)	0.961	0.925	0.998
cpumps Designer/brand name trainers (pov3_18)	0.956	0.914	0.999
cmp3 MP3 player such as an iPod (pov3_19)	0.961	0.923	0.999
(4) Managerial and professional occupations (versus Partly skilled and unskilled manual)			
holiday A holiday away from home for one week a year, not staying with relatives (pov2_02)	0.939	0.894	0.986
visit Visits to friends or family in other parts of the country 4 times a year (pov2_08)	0.921	0.863	0.984
wedding Attending weddings, funerals and other such occasions (pov2_10)	0.929	0.871	0.990
cclubs Children's clubs or activities such as drama or football training (pov4_8)	0.951	0.908	0.996
(5) Tertiary education (versus Primary education)			
car Car (pov1_07)	0.956	0.916	0.998
mobphone Mobile phone (pov1_14)	0.944	0.894	0.996
(6) Catholic (versus Protestant)			
meatfish Meat, fish or vegetarian equivalent every other day (pov1_19)	1.041	1.002	1.082
cleisure Outdoor leisure equipment such as rollerskates, skateboards, footballs etc. (pov3_05)	1.046	1.014	1.078
ctrousers At least 4 pairs of trousers, leggings, jeans or jogging bottoms (pov3_13)	0.954	0.910	0.999
cmoney Pocket money (pov3_16)	0.946	0.904	0.990
(7) Has disability (versus No disability)			
decorate Enough money to keep your home in a decent state of decoration (pov1_01)	1.035	1.003	1.069
twomeal Two meals a day (pov1_22)	1.029	1.003	1.055
table A table, with chairs, at which all the family can eat (pov1_31)	1.043	1.000	1.088
hospital Visiting friends or family in hospital or other institutions (pov2_11)	1.045	1.006	1.086
ccoat A warm winter coat (pov3_07)	1.029	1.003	1.056
cbooks Books at home suitable for their ages (pov3_08)	1.036	1.008	1.064
(8) Has dependants (versus No dependants)			
decorate Enough money to keep your home in a decent state of decoration (pov1_01)	0.950	0.910	0.993
mealfam Friends or family round for a meal or drink at least once a month	1.040	1.001	1.080
cinema Going to the cinema, theatre or music event once a month (pov2_13)	1.059	1.009	1.111
sport Taking part in sport/exercise activities or classes (pov2_14)	1.050	1.004	1.097

Demographic groups	RR	95% CI Lo	95% CI Hi
cbooks Books at home suitable for their ages (pov3_08)	0.956	0.918	0.994
cgarden A garden or outdoor space nearby where they can play safely (pov3_09)	0.961	0.925	0.999

Concluding remarks

A programming error resulted in higher than expected amounts of missing data in the 'sort card' sample of the June 2012 Northern Ireland Omnibus Necessities Module. Some interviewers failed to correctly record all the responses for some respondents. These missing data appear to consist mainly of unrecorded Don't Know responses and no additional biases are evident by demographic group.

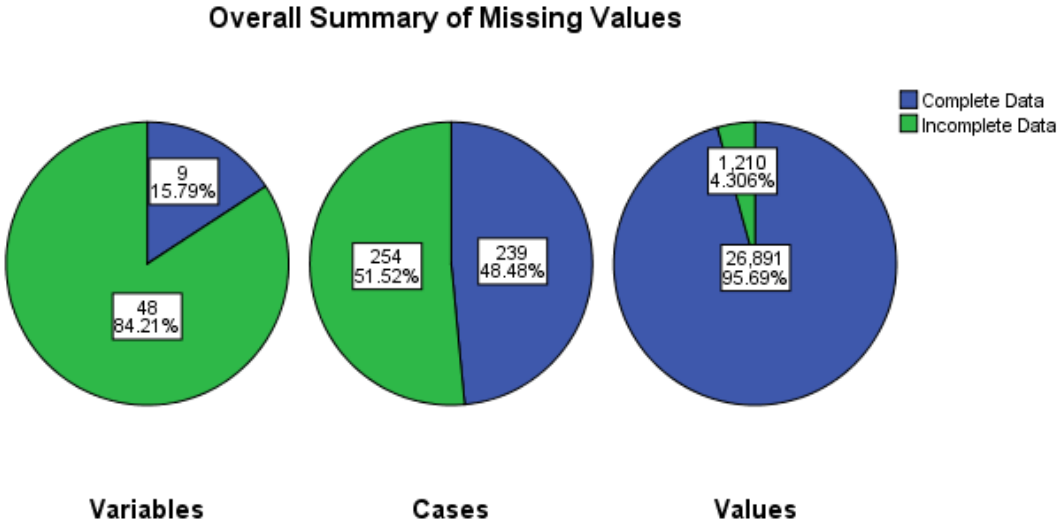
These missing data can be considered to be MAR (Missing At Random) and they are ignorable for analyses of the proportion of respondents considering an item or activity to be a necessity as long as both the missing data and the don't knows are excluded from the numerator and denominator i.e. if the don't knows are set to 'missing'.

Fortunately, this is the normal practice with these kinds of analyses.

APPENDIX: Missing Data Pattern

The missing data analyses results below were produced using the SPSS Multiple Imputation module after excluding sixteen respondents who refused to answer the questions in entire modules i.e. they did not take part in the necessities sort card exercise.

Figure A.1: Adult Items & Activities: Pie Chart Summary of Missing Data



The first pie chart (Figure A.1) shows that only 9 variables had complete data for all respondents and 48 variables had some missing data. The second pie chart shows that just over of respondents (254) had some missing data. Table A.1 shows the missing data frequency for each of the adult items and activities, 43 respondents (8.7%) had a missing card for Dishwasher by contrast only 11 respondents (2.2%) had a missing card for Damp Free Home.

Table A.1: Adult Items and Activities Missing Data Frequencies, NI 2012

Variable Summary[a,b]			
	Missing		Valid N
	N	Percent	
dishwash Dishwasher	43	8.7%	450
pension Regular payments to an occupational or private pension	42	8.5%	451
visit Visit friends or family in other parts of the country four times a year	39	7.9%	454
cinema Going to cinema, theatre or music event once a month	38	7.7%	455
poshfrock An outfit to wear for social or family occasions such as parties, weddings etc	38	7.7%	455
pub Drink out once a fortnight	37	7.5%	456
internet Internet connection at home	35	7.1%	458
sport Taking part in sport or exercise activities or classes	33	6.7%	460
insurance Household contents insurance	33	6.7%	460
furnit Replace any worn out furniture	33	6.7%	460
worship Attend place of worship	31	6.3%	462
mealout Meal out once a month	31	6.3%	462
haircut Hair done or cut regularly	31	6.3%	462
computer Home computer	31	6.3%	462
wedding Attend weddings, funerals and other such occasions	30	6.1%	463
table Table and chairs at which all the family can eat	30	6.1%	463

Variable Summary[a,b]			
	Missing		Valid N
	N	Percent	
money A small amount of money to spend each week on yourself, not on your family	30	6.1%	463
holabrd Holidays abroad once a year	28	5.7%	465
vegfruit Fresh fruit and vegetables every day	28	5.7%	465
roast Roast joint or equivalent once a week	28	5.7%	465
shoes Two pairs of all weather shoes	28	5.7%	465
hospital Visit friends or family in hospital or other institutions	27	5.5%	466
unexcost To be able to pay unexpected costs of £500	26	5.3%	467
mealfam Friends or family around for a meal or drink at least once a month	25	5.1%	468
mobphone Mobile phone	25	5.1%	468
holiday Holiday away from home, not staying with relatives	24	4.9%	469
jobfrock Appropriate clothes for job interviews	24	4.9%	469
savings Regular savings (of at least £20 a month) for rainy days	23	4.7%	470
meatfish Meat, fish or vegetarian equivalent every other day	22	4.5%	471
curtains Curtains or window blinds	21	4.3%	472
phone Telephone	21	4.3%	472
tv Television	21	4.3%	472
clothes Replace worn out clothes with new not second hand clothes	20	4.1%	473
elec Replace or repair broken electrical goods	20	4.1%	473
decorate Enough money to keep your home in a decent state of decoration	20	4.1%	473
presents Presents for family or friends once a year	19	3.9%	474
nightout Going out socially once a fortnight	18	3.7%	475
car Car	17	3.4%	476
celebrat Celebrations on special occasions	15	3.0%	478
hobby Hobby or leisure activity	15	3.0%	478
heating Heating to keep home adequately warm	15	3.0%	478
econact Economic Activity (econact)	15	3.0%	478
empst2 Employment Status (empst2)	15	3.0%	478
dental All recommended dental treatment	14	2.8%	479
twomeal Two meals a day	14	2.8%	479
warmcoat Warm waterproof coat	14	2.8%	479
washing Washing machine	12	2.4%	481
nodamp Damp-free home	11	2.2%	482
a. Maximum number of variables shown: 50			
b. Minimum percentage of missing values for variable to be included: 0.0%			

Figure A.2 shows the pattern of missing data, each row shows a group of cases with the same pattern of missing values. There are a 192 different missing value patterns amongst the 57 variables (shown in the columns), this is a relatively small number of missing value patterns compared with the total possible number (i.e. 2^{57}). The variables have been ordered from lowest number of missing values on the left to highest number on the right

Figure A.3 shows the 10 most frequently occurring patterns of missing data, pattern 1 (no missing data) is by far the most frequent pattern, the other 9 patterns have similar low frequencies.

Figure A.2: Adult Items & Activities: Missing Value Pattern

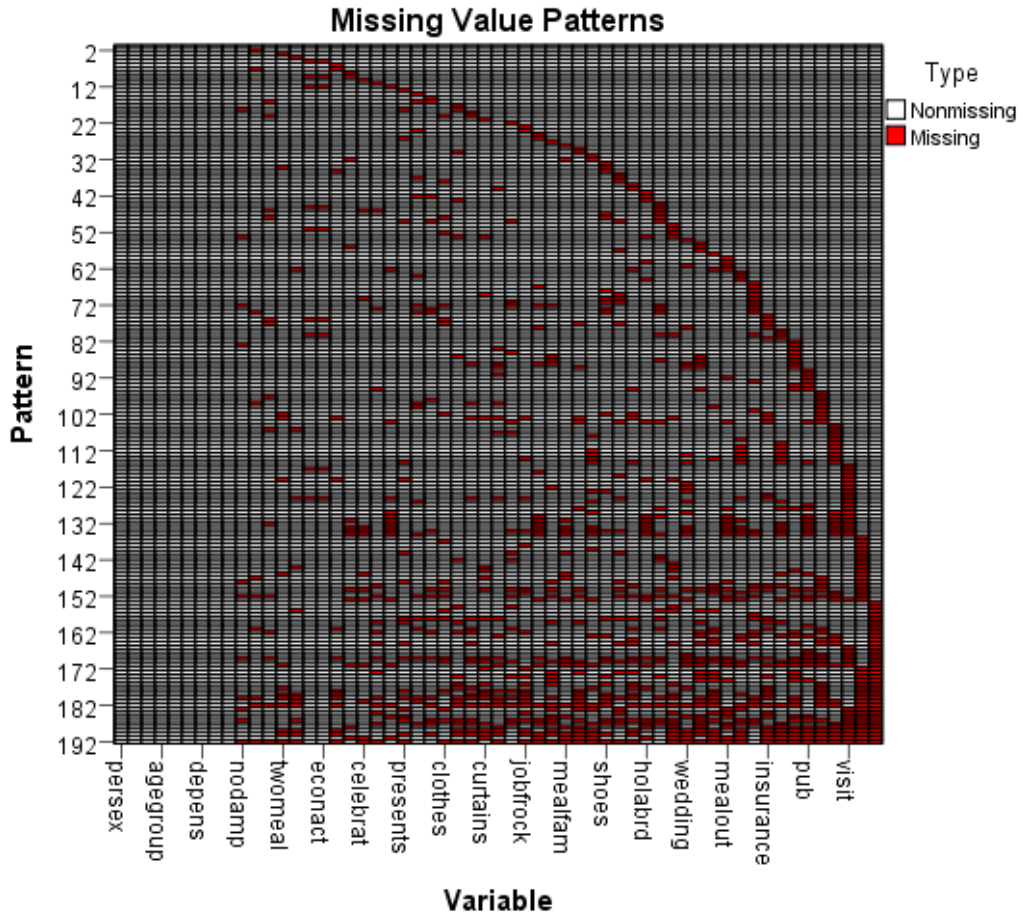
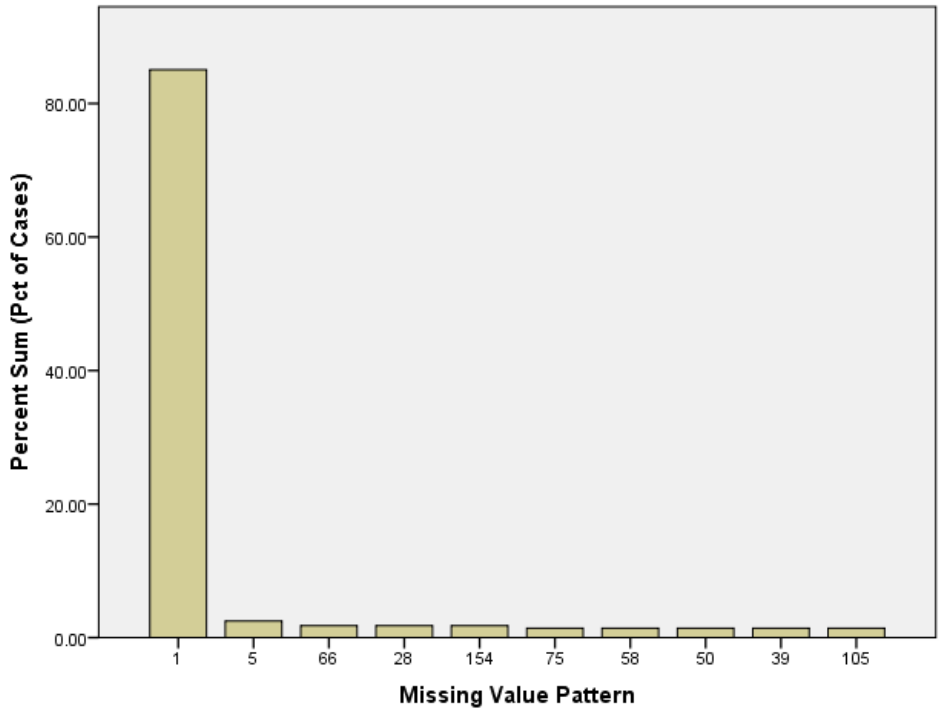


Figure A.3: Adult Items & Activities: Missing Value Bar Chart



The 10 most frequently occurring patterns are shown in the chart.

Figure A.4: Child Items & Activities: Pie Chart Summary of Missing Data

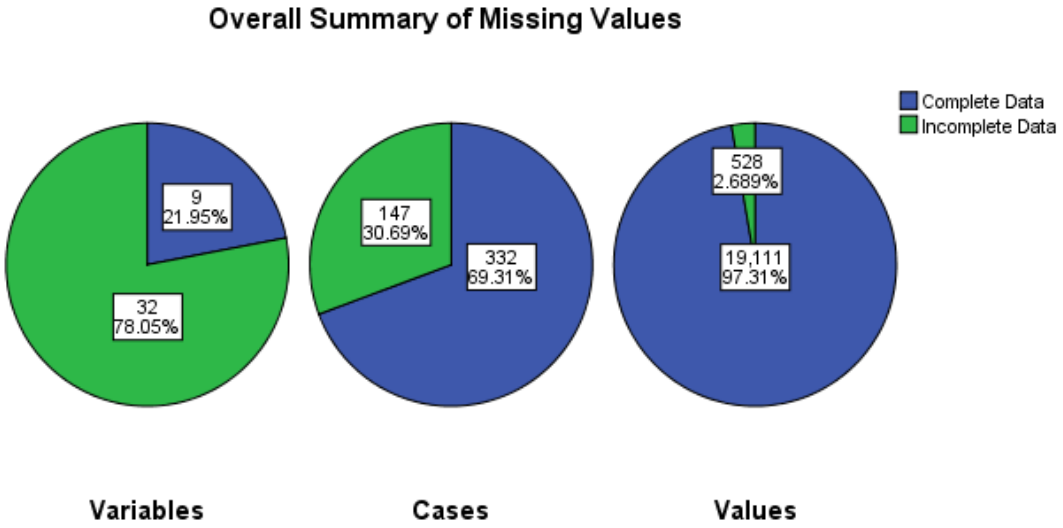


Table A.2: Child Items and Activities Missing Data Frequencies, NI 2012

Variable Summary[a,b]			
	Missing		Valid N
	N	Percent	
cstyle Clothes to fit in with friends (child)	30	6.3%	449
cpumps Designer or brand name trainers (child)	29	6.1%	450
cpc Computer and internet for homework (child)	24	5.0%	455
cmp3 MP3 player such as an Ipod (child)	24	5.0%	455
cBike Bicycle (child)	24	5.0%	455
cmobphone Mobile phone for children aged 11+ (child)	23	4.8%	456
ctrousers At least 4 pairs of trousers, leggings, jeans or jogging bottoms (child)	23	4.8%	456
csave Money to save (child)	21	4.4%	458
cmoney Pocket money (child)	21	4.4%	458
clego Construction toys (like lego, duplo etc) (child)	21	4.4%	458
cmeat Meat, fish or vegetarian equivalent at least once a day (child)	21	4.4%	458
cgames Indoor games suitable for their age (child)	20	4.2%	459
cclothes Some new not second hand clothes (child)	19	4.0%	460
cbooks Books at home suitable for their ages (child)	16	3.3%	463
cgarden Garden or outdoor space to play in safely (child)	15	3.1%	464
cleisure Outdoor leisure equipment, such as roller skates, skateboards, football, etc (child)	15	3.1%	464
econact Economic Activity (econact)	15	3.1%	464
empst2 Employment Status (empst2)	15	3.1%	464
ccoat Warm winter coat (child)	14	2.9%	465
cclubs Childrens clubs or activities such as drama or football training (child)	13	2.7%	466
cschool Going away on a school trip at least once a term (child)	13	2.7%	466
cbedroom Enough bedrooms for every children aged 10+ of a different sex to	13	2.7%	466

Variable Summary[a,b]			
	Missing		Valid N
	N	Percent	
have their own room (child)			
cstudy Suitable place at home to study or do homework (child)	12	2.5%	467
choliday Child holiday away from home for at least 1 week per year (child)	11	2.3%	468
cshoes New properly fitting shoes (child)	11	2.3%	468
cmeal Three meals a day (child)	11	2.3%	468
cplaygrp Toddler group or nursery or play group at least once a week for pre-school aged children (child)	10	2.1%	469
cveg Fresh fruit or veg at least once a day (child)	10	2.1%	469
cfamtrip Day trips with family once a month (child)	9	1.9%	470
csnack Child has friends round for tea or a snack once a fortnight (child)	9	1.9%	470
chobby Child hobby or leisure activity (child)	9	1.9%	470
cceleb Child celebration or special occasions (child)	7	1.5%	472
a. Maximum number of variables shown: 50			
b. Minimum percentage of missing values for variable to be included: 1.0%			

Figure A.5: Child Items & Activities: Missing Value Pattern

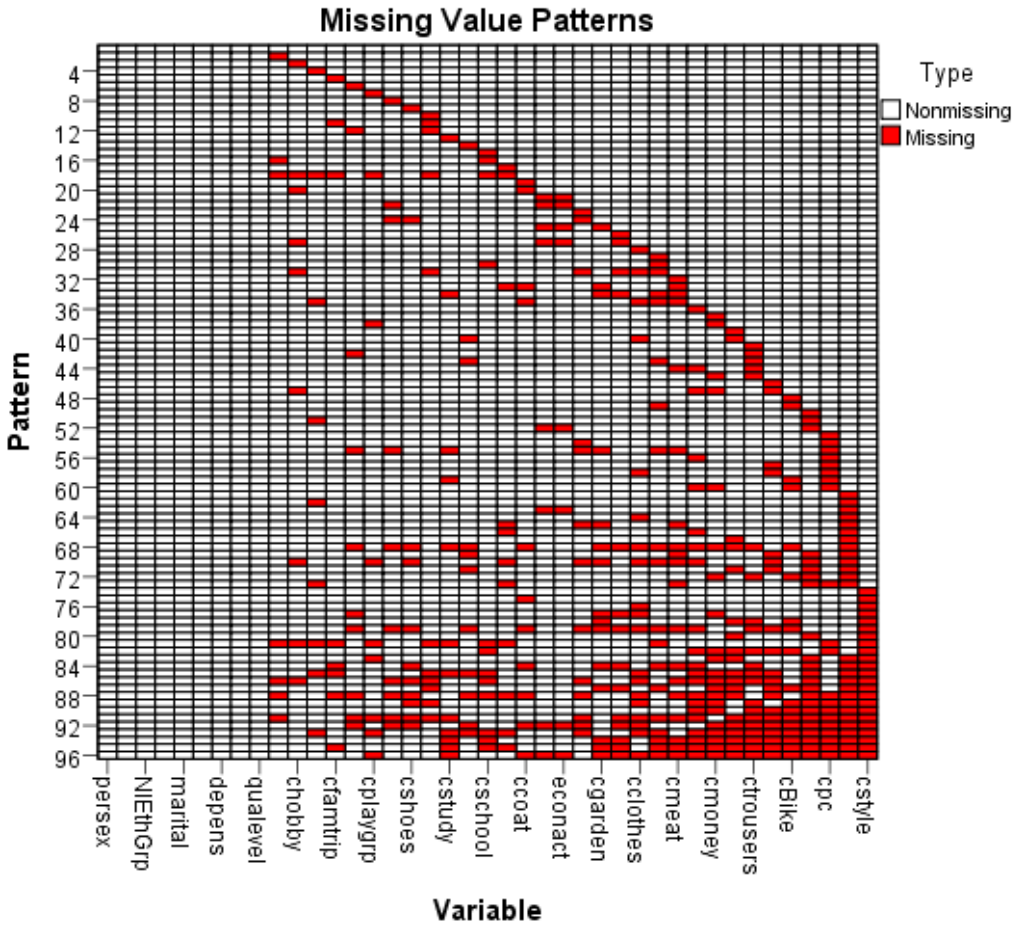
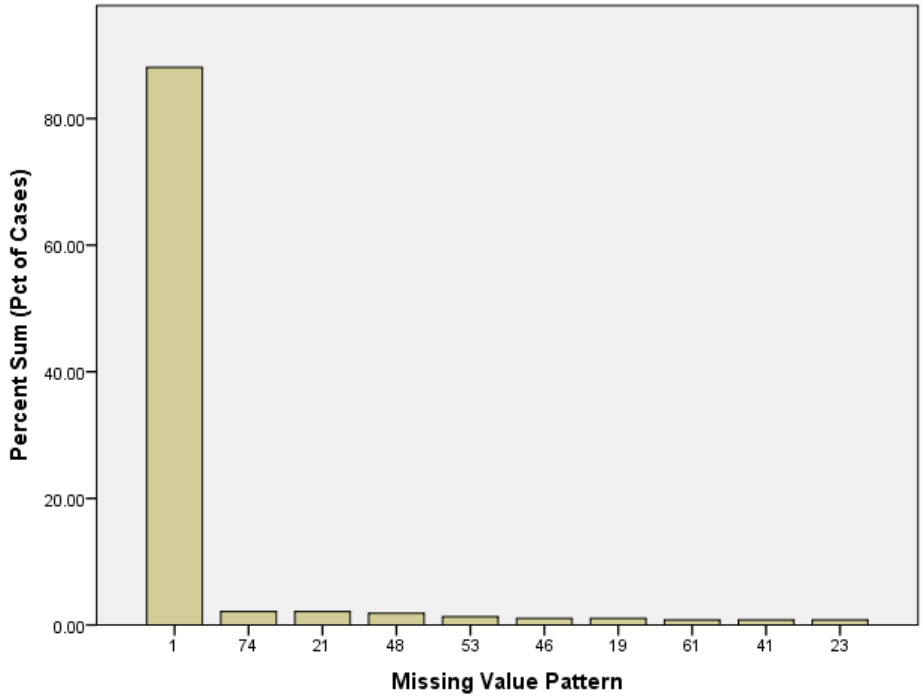


Figure A.6: Child Items & Activities: Missing Value Bar Chart



The 10 most frequently occurring patterns are shown in the chart.

References

Hillyard, P., Kelly, G.P., McLaughlin, E., Patsios, D. and Tomlinson, M. (2003) *Bare Necessities: Poverty and Social Exclusion in Northern Ireland* , Belfast : Democratic Dialogue.

Howell, D. C. (2007) The analysis of missing data. In Outhwaite, W. & Turner, S. *Handbook of Social Science Methodology*. London: Sage.

Little, R.J.A. & Rubin, D.B. (1987) *Statistical analysis with missing data.* New York, Wiley.

Mack, J. and Lansley, S. (1985). *Poor Britain*, London: George Allen & Unwin