

MISSING MIXED MODE: ELEMENTAL STRUCTURES

ESTRUCTURAS BÁSICAS DE LOS VALORES PERDIDOS EN ENCUESTAS CON MODOS MIXTOS

Antonio Alaminos¹

Universidad de Alicante, España

alaminos@ua.es

Recibido: 20/07/2012

Aceptado: 08/10/2012

Abstract

One of the main concerns is the nature of the missing values. Let's consider extremes for simplicity. If missing at random we have not to care about. But if missing shows structures that covariate with substantive variables we have to make decisions. There are, in fact, several options to take. We are speaking about one country, one mode. But if you go cross-cultural (or more precisely, cross-state nations) and mixed modes many questions raise. For example, the simple one. What are we comparing? Reports and books usually go straight into variables distributions and coefficient comparisons. This is possible because the annalist presume "tabula rasa" effect from data collections procedures. But this is not, frequently, the real situation. This paper will expose the mixed missing mode imprint in international surveys. This will help to evaluate how deal with this problem. Also, to consider the real meaning of observed cross-national differences.

Keywords: Non response. Missing values structures. Mode effects. Cross-cultural social survey.

¹ First results of this research were presented at the Large Scale Data Analysis Congress in Cologne. The data 1995 ISSP on National Identity were provided while a research visit at the EUROLAB in Cologne.

Resumen

Una de las principales preocupaciones en la investigación mediante encuestas es la naturaleza de los valores perdidos. En el caso de que sean de ocurrencia aleatoria no generan problemas excesivos. Por el contrario, si los valores perdidos muestran estructuras que covarían con variables relevantes deben tomarse decisiones fundamentales. Eso en el caso de una encuesta en un país. Pero si se trata de una encuesta comparada, efectuada en varios países y además con modos mixtos de recolección de datos, la cuestión deviene bastante más complicada. Entre las preguntas surge la más evidente ¿Qué se está comparando realmente? Con frecuencia los investigadores emplean directamente distribuciones de frecuencia y coeficientes, haciendo “tabula rasa” de otros factores que puedan generar diferencias. Esta investigación muestra como la interacción entre las estructuras de valores perdidos, el modo de recolección de datos y el hecho de la cross-culturalidad es significativa. Se muestra su efecto en los datos.

Palabras clave: No respuesta. Estructuras de valores perdidos. Efectos del modo. Encuesta comparada.

1. MISSING MODES

Usually “Do not know” (DK) and “No answer” (NA) are the two main sources for missing data. Assumptions on the structure of these missing values (missing at random, completely missing at random or not missing at random) are very important for the conclusions of our ordinary analysis of the data, especially when we apply multivariate analysis in the data analysis. There are many different sources when it comes to explaining this underlying structure of missing values, e.g. socio-economic and demographic characteristics of the societies in interaction with sensitive themes. Mode can also be an important source for explaining the structures of missing data. This paper focus on the problem of non-response in cross-cultural analysis. From a wide point of view, any comparison gets a lot of problem from non-response. One of them is the high probability on modelling subpopulation. Gender, age and others sociodemographic variables usually show good association with the probabilities to non-response a question. The impact is different in different countries. That situation can conduct to apply the same model to models different subpopulation in different countries. The final comparison will be clearly misleading.

To control for the imprint of mode in the final information, has been computed the percentage on non-response due to “Do not know” options in all the variables that can be compared in the 1995 ISSP social survey on “national identity”. Not all the variables in this international survey can be directly compared due to “Not Applicable” options or not being asked in some countries, for example. With this constriction we have keep 48 variables in the

analysis. We have build a non-square matrix “countries by variables” containing the percentage on DK for each country for each variable. Depending on the analysis, we have flipped the matrix and sometimes countries enter as variables.

There is a lot of reason for explaining differences in the track of the non-response in different countries. But we deal here with the opposite phenomenon. Many countries look very similar on the way they behave with non-response.

1.1. Shape and levels

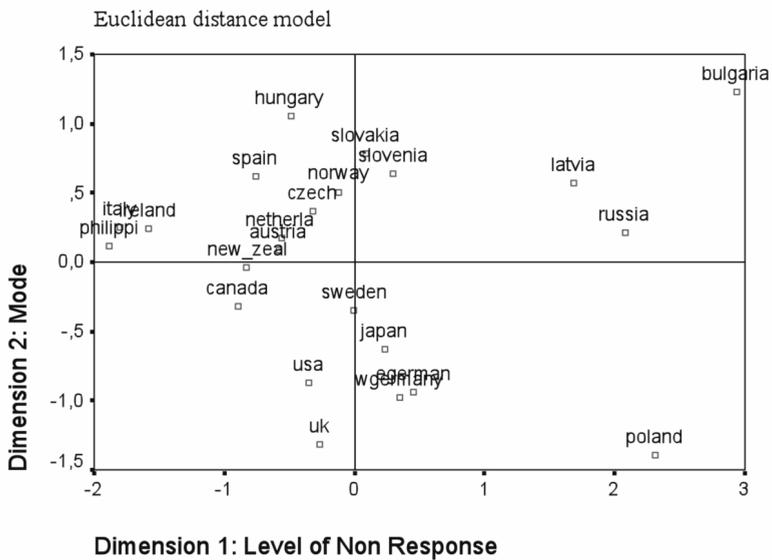
There are two aspects that have to be considered when analysing the profiles on non-response in different countries, shape and level. Shape is the form that the structure on non-response adopts across ordered variables. Level is the mean on non-response taking all the variables into account. If we compare several countries directly, it is possible to identify the high similarities between them.

West and East Germany, Czech and Slovakia, Norway and Sweden, Latvia and Russia or Canada and USA show a very similar pattern in their behaviour toward non-response. These countries have got directly an equivalent shape and level on the pattern of non-response. Their societies react with a higher non-response to the same questions and in a rather similar level of impact. That aspect posits the structural equivalence in the reaction to the measurement on national identity that some societies show. After many years as two separate states, with two different system (political, economical) West and East Germany looks very similar in the way that the two societies expressed through public opinions. The opposite example, with a former state that went into two different, Czech and Slovakia, shows a patter that speaks about a structural equivalence. Something similar can be observed in Latvia and Russia. How a society output in tern on non-response to a questionnaire shows a deep and structural pattern that approaches us to the concept of culture on a broad sense. That idea is easily understood when considering countries like Norway and Sweden or USA and Canada. Departing from this idea we search for that possibility on grouping countries by culture in a multivariate way.

An MDS on states profiles

We have produced a Multidimensional Scaling analysis introducing countries as variables, looking for the matches on the non-response reactions in the different countries. The model fit has been based on Euclidean distance and two dimensions.

Figure 1: derived stimulus configuration



When we look in figure 1 how countries cluster, only Germany behaves as expected. They clustered very close. The main reason for that unexpected fit has to do with the two dimensions detected. The first dimension is essentially level, while the second one is mode. If we consider table 1, we can appreciate the mean level of non-response in the different countries in consideration.

Table 1: means on missing values by country

	N	Mean	Std. Deviation
ITALY	48	2,1119	1,9373
PHILIPPI	48	2,1915	1,6190
IRELAND	48	2,7441	2,3787
SPAIN	48	4,8662	3,5739
CANADA	48	5,1378	3,3813
NEW_ZEAL	48	5,2872	3,0456
HUNGARY	48	5,4447	4,9957
AUSTRIA	48	5,6128	3,9051
NETHERLA	48	5,8371	3,9332
CZECH	48	6,1337	4,4919

	N	Mean	Std. Deviation
USA	48	6,2363	4,4724
SLOVAKIA	48	6,4947	4,9952
NORWAY	48	6,7480	4,5506
SWEDEN	48	7,3013	4,3973
JAPAN	48	7,3497	5,6012
SLOVENIA	48	7,4043	4,9625
UK	48	7,8942	4,2449
WGERMANY	48	8,8647	4,1172
EGERMAN	48	8,9631	4,3817
RUSSIA	48	10,6195	6,5179
LATVIA	48	11,0133	5,7730
POLAND	48	12,8624	7,9950
BULGARIA	48	13,0779	9,0445

Source: ISSP 1995

It is clearly observed that the order of the countries in the first dimension is the order on their level on non-response. Italy, Philippines (2,1), Ireland (2,7), Spain (4,8), Canada (5,1), New Zealand (5,2), Hungary (5,4), Austria (5,6), Netherlands (5,8), Czech (6,1), USA (6,2), Slovakia (6,4), Norway (6,7), Sweden and Japan (7,3), Slovenia (7,4), GB (7,8), West Germany (8,8), East Germany (8,9), Russia (10,8), Latvia (11), Poland (12,8), Bulgaria (13). We can conclude that Dimension one is level on non-response.

When we consider the dimension two we basically recognised the mode of application. Italy, Philippines, Ireland, Spain, Hungary, Austria, Netherlands, Czech, Slovakia, Slovenia, Latvia, Russia, Bulgaria, Japan has “face to face” mode. In the other hand, Canada, New Zealand, USA, Sweden, Norway, Great Britain, West Germany, East Germany and Poland get others modes (phone, mail and mixed modes). Norway and Japan are the only two countries that do not fit the order on that dimension because Japan was “face to face” mode while Norway was done “others” mode.

The dimensional analysis has offered us the presence of two dimensions, mainly “level on non-response” and “mode” of application. Both dimensions are highly correlated. To evaluate this relation we have produced a discriminant analysis considering the “mode” of application as classificatory variable. To

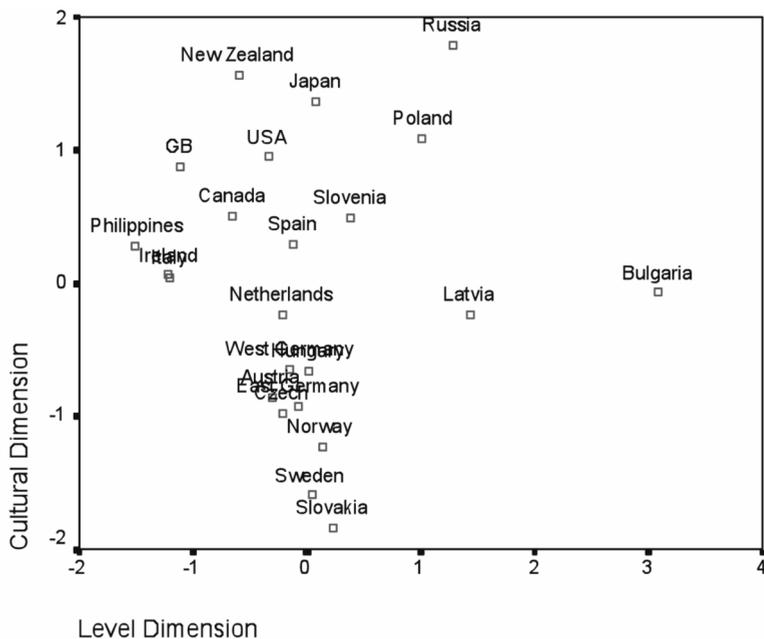
introduce the variability coming from the different groups of question in the questionnaire we have produce factor analysis for the different groups of variables. After this dimensionality reduction, we considered the factors as predictive variables in the discriminant analysis.

If we use the structure on non-response on the different countries to classify the “mode” of application we get a 100% of cases grouped correctly in the ISSP on “national identity”. Japan and Norway are correctly classify on “face to face” and “others” mode. This outcome points out the deep relation in between the structure on non-response and the mode of application. But we still have not solved the main question about cultural similarities in different countries.

1.2. Cultures: Second Order factor analysis

Previously to the discriminant analysis we have translated the individuals items into “first order factors” grouping batteries of question. We have produced a “second order dimensional analysis”, starting from the mentioned “first order factors”. We get three different factors with high Eigen and variance explained. The first dimension is “level” as we tested, the second one cluster on “mode”, and the third one cluster “cultures” in a broad sense.

Figure 2: graph of dimension 3 against dimension 1



In figure 2 it's possible to appreciate how New Zealand, Japan, USA, Great Britain, Canada, Slovenia, Philippines, Spain, Italy, Ireland form a wider cluster. Netherlands, Hungary, West Germany, Austria, East Germany, Czech Republic, Norway, Sweden, Slovakia another; Russia, Poland, Latvia and Bulgaria a third one. It is possible to put different brackets (Religions for example) but it looks as South-West culture, North-Central Europe and Central-East Europe.

When considering how societies behave on non-response, there are three different dimensions that affect the analysis cross-culturally. The first dimension is the "level" or mean on the global non-response to a questionnaire (mainly the problem of analysing subpopulation due to social desirability impact). The second dimension to be taken into account is "mode" or the method of questionnaire application. The third dimension has to do with "culture" in a broad sense, which offers different structures in the sequences of non-response across countries.

We may conclude from an empirical point of view that: when dealing with social surveys cross-culturally, it is essential to control the effect of "mode" (application), to control the differences on "level" (non-response), and control the effect of the "culture" (that rule the pattern). These three dimensions produce an effect that contributes to disturb the comparisons. Specially, the interaction between mixed modes and cultures has to be carefully controlled.

ANTONIO ALAMINOS es Doctor en Sociología por la Universidad Complutense de Madrid. En la actualidad es Catedrático de Sociología en el Departamento de Sociología II de la Universidad de Alicante e investigador del Instituto Interuniversitario de Desarrollo Social y Paz. Sus principales líneas de investigación son integración europea, migración y aculturación, socialización, comportamiento político, desarrollo de escalas e indicadores, y métodos de investigación.