UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-36
Statistics and Econometrics Series 28
November 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

PROFILE IDENTIFICATION VIA WEIGHTED RELATED METRIC SCALING: AN APPLICATION TO DEPENDENT SPANISH CHILDREN

Irene Albarrán[(1)], Pablo Alonso[(2)] and Aurea Grané[(1)]

Abstract:

Disability and dependency (lack of autonomy in performing common everyday actions) affect health status and quality of life, therefore they are significant public health issues. The main purpose of this study is to establish the existing relationship among different variables (continuous, categorical and binary) referred to children between 3 and 6 years old and their functional dependence in basic activities of daily living. We combine different types of information via weighted related metric scaling to obtain homogeneous profiles for dependent Spanish children. The redundant information between groups of variables is modeled with an interaction parameter that can be optimized according to several criteria. In this paper, the goal is to obtain maximum explained variability in an Euclidean configuration. Data comes from the Survey about Disabilities, Personal Autonomy and Dependence Situations, EDAD 2008, (Spanish National Institute of Statistics, 2008).

Keywords: ADL, Disability, Mixed-type Data, Public Health, Related Metric Scaling.

AMS subject classification: 62-07, 62-09, 62H20, 62H99, 62P05.

Authors' address: [(1)] Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain. (2) Statistics Department, Universidad de Alcalá, Pza. San Diego, s/n - 28801 Alcalá de Henares (Madrid).
E-mail: A. Grané, aurea.grane@uc3m.es (Corresponding author), I. Albarrán, irene.albarran@uc3m.es, P. Alonso, pablo.alonso@uah.es .
This work has been partially supported by Spanish grant MTM2010-17323 (Spanish Ministry of Science and Innovation).

Profile identification via weighted related metric scaling: An application to dependent Spanish children

Irene Albarrán[(1)] Pablo Alonso[(2)] Aurea Grané[(1)]

(1) Statistics Department. Universidad Carlos III de Madrid.
(2) Statistics Department. Universidad de Alcalá.

Abstract

Disability and dependency (lack of autonomy in performing common everyday actions) affect health status and quality of life, therefore they are significant public health issues. The main purpose of this study is to establish the existing relationship among different variables (continuous, categorical and binary) referred to children between 3 and 6 years old and their functional dependence in basic activities of daily living. We combine different types of information via weighted related metric scaling to obtain homogeneous profiles for dependent Spanish children. The redundant information between groups of variables is modelled with an interaction parameter that can be optimized according to several criteria. In this paper, the goal is to obtain maximum explained variability in an Euclidean configuration. Data comes from the Survey about Disabilities, Personal Autonomy and Dependence Situations, EDAD 2008, (Spanish National Institute of Statistics, 2008).

Keywords: ADL, Disability, Mixed-type Data, Public Health, Related Metric Scaling

AMS Subject Classification: 62-07, 62-09, 62H20, 62H99, 62P05.

1 Introduction

Living is not always an easy task, specially if the physical or mental conditions are not in their fullness. This is the case when people are not able to carry out certain activities that are common in our daily life because they suffer some disabilities. This situation refers to the negative aspects of the interaction between the individual and the environment, i.e., deficits, limitations in the activity and restrictions in his/her social participation (WHO 2001b). This set of obstacles turns into a tougher situation when a third person is required to do these activities, which is the case of dependency. Disability has traditionally been a marginalized concern of public health and has usually been viewed as a failure of primary prevention. However, disparities in

Financial support from research project MTM2010-17323 by the Spanish Ministry of Science and Innovation.
Authors' address: (1) Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain. (2) Statistics Department, Universidad de Alcalá, Plaza de la Victoria 2, 28802 Alcalá de Henares, Spain. E–mails: I. Albarrán, irene.albarran@uc3m.es, P. Alonso, pablo.alonsog@uah.es, A. Grané, aurea.grane@uc3m.es
Corresponding author: A. Grané; Date: December 1, 2011.

health behaviours, health access, and health status between people with and without disabilities suggest that opportunities exist for public health to engage people with disabilities to improve their overall health (Crews and Lollar 2008). Disability is a large public health problem in the main developed countries. It has been analysed in the US (Pope and Tarlov 1991) and in the Western European countries (Haveman and Wolfe 2000). The International Classification of Functioning, Disability and Health (ICF)(WHO 2001b) tries to establish a consensus in its understanding, by establishing a difference between the basic activities of the daily life and the instrumental activities of the daily life. The first ones are defined as those essential activities for an independent life. There are many ways to define what *dependency* is. One of the most accepted is that included in Resolution R(98) of the Council of Europe that defines it as "such state in which people, whom for reason connected to the lack or loss of physical, mental or intellectual autonomy, require assistance and/or extensive help in order to carry out common everyday actions".

This general definition has been translated to national legislations in an heterogeneous way (Kamette 2011). In fact, it may happen that a man/woman can be considered dependant in a country but not in another. For instance, a 58 year-old dependent man in Spain or in Germany will not be considered dependant in France because in the French system it is required to be over 60 years old. Or, in a another way, the intensity of the legally recognized dependence may change amongst countries (see Albarrán, Alonso, and Bolancé 2009). If we consider the Spanish case, the definition of dependency is that included in article 2 of Act 39/2006, of 14^{th} December, on the Promotion, Personal Autonomy and care for Dependent persons. It is defined as a "permanent state in which persons that for reasons derived from age, illness or disability and linked to the lack or loss of physical, mental, intellectual or sensorial autonomy require the care of another person/other people or significant help in order to perform basic activities of daily living or, in the case of people with mental disabilities or illness, other support for personal autonomy".

Another key element is the *assessment of dependency*. This question is usually solved using specific scales that take into account the disabilities suffered by the person jointly with their intensity. It is not easy to measure this last matter, being one of the most extended solutions to evaluate the time dedicated by a third person to help him/her to do certain activities, such as dressing or eating by him/herself. This is how it is assessed in several national systems such as those currently in force in Spain or in Germany. The evaluation of dependency in Spain is ruled by the Royal Decree 504/2007, of 20^{th} April, that passes the scale for assessment of the situations of dependency set by Act 39/2006. According to it, the scale goes from 0 to 100 points and at least 25 points are needed to acknow the entitlement to the benefits of the System. In Table 1 we show the dependency graduation following Spanish legislation. According to the scale value reached by an individual, Act 39/2006 establishes a minimum level of protection, which is defined and financially guaranteed by the General State Administration. Dependent persons shall be entitled to access under equal conditions to the benefits and services foreseen in this Act, according to the terms laid down in it. For instance, regarding dependent children, some family benefits, social service benefits in re-education and rehabilitation can be obtained. This scale is used in all cases when individuals are over 3 years old. Nevertheless, when the individuals are under 6, the International Classification of Functioning (ICF) mentioned before

Table 1: Dependency graduation following Spanish legislation

Dependency	Degree	Level	Scale values	Dependency	Degree	Level	Scale values
Non dependant	-	-	$[0, 25)$	Severe	II	1	$[50, 65)$
					II	2	$[65, 75)$
Moderate	I	1	$[25, 40)$	Major	III	1	$[75, 90)$
	I	2	$[40, 50)$		III	2	$[90, 100]$

Moderate dependency	The person needs help in order to perform various basic ADL[(1)] at least once a day or the person needs intermittent or limited support for his/her personal autonomy.
Severe dependency	The person needs help in order to perform various basic ADL two or three times a day, but he/she does not want the permanent support of a carer or he/she needs extensive support for his/her personal autonomy.
Major dependency	The person needs help in order to perform various basic ADL several times a day or he/she needs the indispensable and continuous support of another person or he/she needs generalised support for his/her personal autonomy.

[(1)] ADL stands for Activities of Daily Living.

is replaced by its version for children and youth (ICF-CY) (WHO 2001a). This version is based on the same model as ICF with added content adapted to these groups of people. ICF is meant to provide a common language to professionals and other stakeholders involved in facilitating functioning for persons with body impairments and activity limitations. Besides, in this case the term limitation is used instead of disability. However, the federal Maternal and Child Health Bureau (MCHB) defines *children with special health care needs* as those for a chronic physical, developmental, behavioral, or emotional condition and who also require health and related services of a type or amount beyond that required by children generally (see USDHHS 2004). Dependence is the main impact factor on health and quality of life (Millán-Calenti 2006). There are many studies about dependence when people is over 65 (see Grammenos 2003, Lafortune and Balestat 2007, Giannakouris 2009). However, concerning children, there is a lack of research using internationally accepted measurements. Benavente and Pfeiffer (2002) pointed out that children with disabilities is a large area to be covered. In particular, our contribution is to study those children with specific disabilities linked to some limitations (see Annex I) who are considered dependant according to the Spanish legislation. Becoming dependant during the first stages of life implies that the individual is going to need special cares during the rest of his/her life (Claeson and Waldman 2000 and Hauser-Cram, Warfield, and Shonkoff 2001). Nowadays, thanks to the combined efforts of medicine, public health and policy, children with chronic conditions or disability live to adulthood, often with a life span similar to the general population (Cannell, Brumback, Bouldin, Hess, Wood, Sloyer, Reiss, and Andresen 2011). For this reason, it seems quite necessary to know as best as possible how the Spanish child population is living in these circumstances, that is, we are interested in establishing homogeneous profiles. Each of these groups will have different necessities (for instance, medical, psychological or social cares) with different economic consequences. The aim of this paper is to classify into different groups

the dependent Spanish population between 3 and 6 years old. These groups will be created depending on certain inherent characteristics such as sex, suffered limitation and its severity or weekly hours of personal assistance. The statistical information comes from the Survey about Disabilities, Personal Autonomy and Dependence Situations, 2008 (EDAD 2008, according to its Spanish acronym). This question, which is crucial in actuarial science, has been usually solved by classical segmentation techniques, such as k-means (see, for instance, Anderberg 1973 and Morgan and Ray 1995). Applications to class rating definitions may be found in Loimaranta, Jacobsson, and Lonka (1980), Campbell (1986), Boj, Claramunt, and Fortiana (2001) and Boj, Claramunt, and Fortiana (2004). However, technical relevant problems arise because of (i) the sampling design and (ii) the nature of the variables observed. Indeed, in EDAD 2008 a two-stage sampling was conducted by INE, leading to individuals that represent population groups of different sizes. We refer to this situation as a weighted context, in contrast to the classical iid sampling. Regarding the nature of the variables, data collected through questionnaires are often of mixed-type obtained as measures of variables at different levels, e.g. quantitative, multi-scale categorical and binary variables. The repertory of statistical techniques suitable for this mixed-type data is scarce. Among them, multidimensional scaling appears to be one of the most flexible techniques dealing with mixed-type data. Bearing in mind the setting of homogeneous profiles, it becomes crucial the definition of a proper distance (or dissimilarity) function among individuals. The well-known Gower's general similarity coefficient (see Gower 1971, Cox and Cox 2000) considers mixtures of quantitative, multi-scale categorical and binary variables. Nevertheless, the additive treatment of the variables of Gower's based similarity coefficients results in a lack of consideration of the association between variables. Grané and Romera (2009) proposed to construct a joint metric via related metric scaling (Cuadras and Fortiana 1998) from three different distance matrices computed on quantitative, multi-scale categorical and binary variables, respectively. Through a case study these authors show the potential of this technique, which overperforms the classical Gower's metric, and study the sensitivity and robustness of their proposal through crossvalidation procedures. This paper extends their proposal in two directions. Firstly, we consider related metric scaling in the weighted context, in the sense that each individual can represent a group of individuals, and secondly we model the redundant information between groups of variables through an interaction parameter, that can be optimized according to several criteria (see also Esteve 2003). For example, in the case study of data coming from EDAD 2008, since we are interested in obtaining homogeneous profiles for dependent Spanish children, the interaction parameter is achieved by imposing maximum explained variability in the Euclidean configuration.

Searching for profiles of homogeneous dependent Spanish children we find out that the Spanish scale is not sufficient in order to measure properly the severity of the situation of dependency, since the time devoted to care dependent children and the intensity of the dependency is not directly related. This is quite relevant, since the scale value reached by an individual allows him/her to access to the benefits entitled by the Spanish Act 39/2006. Finally, we have observed that the definition and classification that the International Classification of Functioning for children and youth (WHO 2001a) establishes can be refined and complemented according to USDHHS (2004).

The rest of the paper is organised as follows: the Spanish database is described in

Section 2. Section 3 is devoted to multidimensional scaling methodology for mixed-type data in the weighted context, where the new proposal of weighted related metric scaling is introduced. An application of this new methodology can be found in Section 4, where we obtain different homogenous profiles for dependent Spanish children using data coming from EDAD 2008. Finally, we conclude in Section 5.

2 Database used in the analysis

Three surveys about disability have been undertaken by INE (Spanish National Institute of Statistics) during the last 25 years in Spain. The first one, elaborated in 1986, was the Survey about Disabilities, Impairments and Handicaps (EDDM 1986, according its Spanish acronym). The next one, the Survey about Disabilities, Impairments and Health Status (EDDES 1999, according its Spanish acronym), was prepared using data of 1999. Finally, the last one was the Survey about Disabilities, Personal Autonomy and Dependence Situations (EDAD 2008, according its Spanish acronym). Although all of them talk about disabilities, it is impossible to track this phenomenon in a homogeneous way along the years because the definition of that concept has been changing through the years depending on the classification used to prepare the survey.

2.1 Recent disability survey in Spain: EDAD 2008

In order to provide reliable estimates at the national level, the survey was performed around the country using sampling. In particular, a two-stage sampling was performed, stratified and proportional to the size of the Spanish autonomous regions (with stratified sampling distribution proportional to population size in stratum, within each Spanish province). See INE (2010) for more details on the sampling methodology.

EDAD 2008 gives information about people with disabilities that were living either in a particular home or in institutions. In the first case, the survey was prepared interviewing 260,000 people who were living in 96,000 different houses whereas for institutionalized people, 11,000 people in 800 centers were asked about their situation. This survey is based on the concept of self perceived disability, in accordance with the recommendations of the World Health Organization. So, the target people is identified through a set of questions about the possible difficulties they can find in doing some specific activities. Despite its drawbacks, the main advantage of this strategy is that it is focused in the daily activities of the individuals and the problems they may have while doing them, with no consideration of medical matters. That is, it puts the attention of both interviewer and interviewed in functional affairs since they are key aspects when talking about disability (Jiménez and Huete 2010).

According to EDAD 2008, there are more than 4.1 million Spanish people suffering at least one kind of disability, 3.85 million out of them living with their relatives or in their own homes, whereas the remaining 0.27 millions are in specialized centers. Although the global prevalence rate is 9.1%, in the case of people living at home, this rate is lower than that for people living in institutions (8.5% and 17.7%, respectively). Disability is mainly related to two main variables: sex and age. Until 45 years old, the male prevalence is greater than the female one. After that age, the relative incidence

is greater for women. In general terms, more than 50% people with this problem are at least 65 years old, being most of them women. Figures about people affected that are living at home can be seen in Table 2.

Table 2: People with disability living at home: number and prevalence rate

	Number (000)			Prevalence rate (%)		
Ages (years)	Total	Men	Women	Total	Men	Women
Under 6	60.4	36.4	24.0	2.2%	2.5%	1.8%
Between 6 and 44	608.1	345.1	263.0	2.5%	2.8%	2.3%
Between 45 and 64	951.8	409.0	542.8	8.7%	7.6%	9.8%
65 or more	2,227.5	756.7	1,470.8	30.3%	24.1%	34.9%
Total	3,847.8	1,547.2	2,300.6	8.5%	6.9%	10.1%

Source: own elaboration using EDAD 2008

Despite the fact that the survey includes the term "dependence" into its denomination, the questionnaire does not consider questions on this topic. In fact, if we looked for the number of dependants reflected in the survey, we would not be able to know how many individuals would be in this situation. So, the only way to answer this question is trying to apply as best as possible both the definition incorporated in article 2 of Act 39/2006 and the assessment scale regulated by Royal Decree 504/2007. Hence, the result is an estimation.
Besides this problem, there is another aspect that makes the study of this contingency in children even more difficult. It must be considered that the analysis of this phenomenon for population until 6 years old has to be done using the ICF-CY Classification, where the concept of disability is replaced by that of limitation, because children are dependant by themselves. For instance, it has no sense to talk about self care in children. Moreover, there are other limitations that can only be seen during the growth of a child, i.e., some difficulties in speaking. Therefore, it is no surprising that the proportion of children with limitations increases with age (Grupo de Atención Temprana 2000). In addition, children in the earlier ages (0-2 years old) are assessed with an special scale whose concepts are not reflected in EDAD 2008. This is the reason why this paper is focused on children between 3 and 6 years old. In a strict sense, the considered ages are those between 36 and 71 months old.

2.2 Description of the data set

After having filtered the information with the definition of dependency included in Act 39/2006, the number of records to be analysed is 84. Taking into account the sampling methodology, the Spanish National Institute of Statistics estimates that the number of dependent Spanish children with possibilities of receiving public aid is 13,296. Their number and prevalence rate by age and gender are shown in Table 3.
In Table 4 we briefly describe the twenty-four mixed-type variables considered in the analysis. They consist of three continuous variables such as the age of the child (in months), the scale value reached by the child and weekly hours of attention, three multi-state categorical variables such as some information about the respondent, the type of received aids and the severity of limitations to perform activities of daily

Table 3: Dependent children: number and prevalence rate by age and gender

	Number			Prevalence rate (%)		
Ages (months)	Total	Boys	Girls	Total	Boys	Girls
Between 36 and 47	2,751	1,702	1,049	0.6%	0.7%	0.5%
Between 48 and 59	5,621	3,473	2,148	1.2%	1.5%	0.9%
Between 60 and 71	4,923	3,272	1,652	1.1%	1.4%	0.8%

Source: own elaboration using EDAD 2008

living (ADL) and, finally, eighteen binary variables such as sex and several limitations described in the Annex.

Table 4: Variables included in the analysis with its possible values

Type	Description	Values/categories (% frequency distribution)	Type	Description	Values/categories (% frequency distribution)
B[(1)]	sex	male(66.6%), female (33.3%)	B	lim 15	yes (7.9%), no(92.1%)
B	lim 1[(2)]	yes (13.1%), no(86.9%)	B	lim 16	yes (70.1%), no(29.9%)
B	lim 2	yes (21.5%), no(78.5%)	B	lim 17	yes (76.5%), no(25.5%)
B	lim 3	yes (26.6%), no(73.4%)	B	lim 18	yes (79.3%), no(20.7%)
B	lim 5	yes (7.6%), no(92.4%)	C	age (months)	from 36 to 71
B	lim 6	yes (2.0%), no(98.0%)	C	scale	from 0 to 100
B	lim 7	yes (14.4%), no(85.6%)	C	hours-week	from 0 to 168
B	lim 8	yes (24.5%), no(75.5%)	CT	inf-relac	parents (92.8%), tutor (2.4%),
B	lim 9	yes (36.1%), no(63.9%)		(respondent)	grandparents (4.8%)
B	lim 10	yes (17.1%), no(82.9%)	CT	B-2	only personal assistance (67.1%),
B	lim 11	yes (80.9%), no(19.1%)		(received aids)	personal assistance and aids (32.9%),
B	lim 12	yes (7.6%), no(92.4%)			only technical aids (0.0%)
B	lim 13	yes (33.4%), no(66.6%)	CT	B-5 (severity	moderate(69.9%), severe (19.2%),
B	lim 14	yes (30.2%), no(69.8%)		of limitations to perform ADL)	cannot perform ADL[(3)] (10.9%)

[(1)] B=binary, C=continuos, CT=categorical. [(2)] See Annex I for the definition of limitations lim 1 to lim 18.
[(3)] ADL stands for Activities of Daily Living.

3 Weighted Multidimensional Scaling for mixed-type data

Multidimensional Scaling (MDS) is a multivariate technique closely related to Principal Component Analysis (PCA) and Correspondence Analysis (CA), well-known techniques and widely used by applied researchers. The objective of these techniques is the description and the pictorial representation of a data set. The information provided by the data set may be a matrix of observations corresponding to a set of continuous variables, which is the case of PCA, a contingency table obtained from the classification of a set of objects according to categorical variables, which is the case of CA, and for the MDS the data set is a square matrix of dissimilarities between a set of objects. The main advantage of MDS is that it is able to cope with variables of any type (binary, categorical, numerical, functional, ...) or even a mixture of them, since using a proper "distance" function one can obtain the matrix of dissimilarities

between the set of objects (see Ramsay 1980). In particular, the purpose of MDS is to construct a set of points in a Euclidean space whose interdistances are either equal (metric or classical MDS) or approximately equal (nonmetric MDS) to those in a given matrix of dissimilarities, in such a way that the interpoint distances approximate the interobject dissimilarities as closely as possible. That is, given a $n \times n$ matrix $\mathbf{\Delta}$, containing the squares of dissimilarities between n objects, the goal is to obtain a n-point configuration onto orthogonal axes (called Euclidean configuration/map or MDS configuration), so that the L^2–distances between the coordinates of these n points coincide with the corresponding entries in $\mathbf{\Delta}$. These coordinates are called a metric scaling representation of $\mathbf{\Delta}$. Various possible measures of approximation between interpoint distances and interobject dissimilarities can be used, each resulting in a different MDS configuration. In this work, these coordinates are obtained via spectral decomposition. General context references are Borg and Groenen (2005), Cox and Cox (2000) and Krzanowski and Marriott (1994) as well as Gower and Hand (1996).

In the following we review the extension of classical MDS concepts to the weighted context, derived by Boj, Claramunt, and Fortiana (2001). Recall that in the weighted context each individual can represent a population group of different size.

Given n p-dimensional vectors $\{\mathbf{z}_i, 1 \leq i \leq n\}$ containing the information of the n different individuals we compute a squared distances matrix $\mathbf{\Delta}$, with entries $\delta^2(\mathbf{z}_i, \mathbf{z}_j)$, for $1 \leq i, j \leq n$. Since this information can be either of qualitative or quantitative nature, or both, it is crucial the adequacy of the dissimilarity function used in the computation of $\mathbf{\Delta}$. Additionally, we have $\mathbf{w} = (w_1, \ldots, w_n)'$ a vector of weights, such that $w_i > 0$, for $i = 1, \ldots, n$, and $\mathbf{1}'\mathbf{w} = 1$, where $\mathbf{1}$ is the $n \times 1$ vector of ones.

Suppose that we are interested in obtaining a metric scaling representation of $\mathbf{\Delta}$, provided that $\mathbf{\Delta}$ satisfies the Euclidean requirement. Given $\mathbf{w}$, define $\mathbf{D_w} = diag(\mathbf{w})$, a $n \times n$ diagonal matrix whose diagonal is the vector of weights, and $\mathbf{K_w} = \mathbf{1w'}$, then $\mathbf{J_w} = \mathbf{I} - \mathbf{K_w}$ is the $\mathbf{w}$-centering matrix, which is an orthogonal projector with respect to $\mathbf{D_w}$, idempotent and self-adjoint with respect to $\mathbf{D_w}$. Then, the doubly $\mathbf{w}$-centered inner-product matrix is

$$\mathbf{G_w} = -\frac{1}{2}\mathbf{J_w}\,\mathbf{\Delta}\,\mathbf{J_w}$$

and

$$\mathbf{F_w} = \mathbf{D_w}^{1/2}\,\mathbf{G_w}\,\mathbf{D_w}^{1/2} \tag{1}$$

is the standardized inner-product matrix. The Euclidean requirement is equivalent to the positive semi-definiteness of $\mathbf{G_w}$, hence to the existence of an $\mathbf{X_w}$ such that $\mathbf{G_w} = \mathbf{X_w}\,\mathbf{X_w'}$, called in the weighted context a $\mathbf{w}$-centered Euclidean representation of $\mathbf{\Delta}$, meaning that $\mathbf{w}'\,\mathbf{X_w} = \mathbf{0}$ and that the squared Euclidean interdistances between the rows of $\mathbf{X_w}$ coincide with the corresponding entries in $\mathbf{\Delta}$. [1]

This matrix $\mathbf{X_w}$ is the $\mathbf{w}$-weighted metric scaling representation of $\mathbf{\Delta}$, which is obtained through the spectral decomposition of (1) as

$$\mathbf{X_w} = \mathbf{D_w}^{-1/2}\,\mathbf{U}\,\mathbf{\Lambda}, \tag{2}$$

[1] If some of the eigenvalues of $\mathbf{G_w}$ are negative, then $\mathbf{\Delta}$ does not admit an Euclidean configuration, which means that some of the axes in the representation are imaginary. In this case, a possible solution (still valid in the weighted context, since $\mathbf{w}$ is an eigenvector of $\mathbf{G_w}$ of 0 eigenvalue) is to consider the transformation $\tilde{\mathbf{\Delta}} = \mathbf{\Delta} + c\,(\mathbf{1}_n\,\mathbf{1}_n' - \mathbf{I}_n)$, where $c \geq 2|\lambda|$ and λ is the negative eigenvalue of maximum module, which assures an Euclidean configuration for $\tilde{\mathbf{\Delta}}$.

where $\mathbf{\Lambda}^2$ is a diagonal matrix containing the eigenvalues of $\mathbf{F_w}$, ordered in decreasing order, and $\mathbf{U}$ is the matrix whose columns are the corresponding eigenvectors. The rows of $\mathbf{X_w}$ contain the principal coordinates of the n individuals and its columns are the principal axes of this representation.
In the following we describe two ways of obtaining a $\mathbf{w}$-weighted metric scaling representation of $\mathbf{\Delta}$ from p continuous and categorical variables measured on a set of n individuals with a weight vector $\mathbf{w}$. The first one is the classical approach and proceeds by computing Gower's general similarity coefficient in the weighted context, whereas the second one is called weighted related metric scaling and extends the proposal of Cuadras and Fortiana (1998) to the weighted context.

3.1 Classical approach: Gower's general similarity coefficient

After a review of the specialized literature, we found that the most popular similarity measure in the context of mixed-type data is the well-known Gower's general similarity coefficient (see Gower 1971), which for two p-dimensional vectors $\mathbf{z}_i$ and $\mathbf{z}_j$ is equal to

$$s_{ij} = \frac{\sum_{h=1}^{p_1} (1 - |z_{ih} - z_{jh}|/R_h) + a + \alpha}{p_1 + (p_2 - d) + p_3}, \tag{3}$$

where $p = p_1 + p_2 + p_3$, p_1 is the number of continuous variables, a and d are the number of positive and negative matches, respectively, for the p_2 binary variables, α is the number of matches for the p_3 multi-state categorical variables, and R_h is the range of the h-th continuous variable. The entries of matrix $\mathbf{\Delta}$ are computed as

$$\delta^2(\mathbf{z}_i, \mathbf{z}_j) = 1 - s_{ij}. \tag{4}$$

Gower (1971) proved that (4) satisfies the Euclidean requirement.

3.2 Weighted Related Metric Scaling

Like all distance functions satisfying additivity with respect to variables, the distance based on Gower's general similarity coefficient implicitly ignores any association (e.g. correlation) between variables (Gower 1992, Krzanowski 1994). Alternative metrics have been proposed in the literature to overcome that problem, among them, we decided to extend *Related Metric Scaling* (Cuadras and Fortiana 1998) to the weighted context. Related metric scaling is a multivariate technique that allows to obtain a unique representation of a set of individuals from several distance matrices computed on the same set of individuals. The method is based on the construction of a joint metric that satisfies several axioms related to the property of identifying and discarding redundant or repeated information.
Given a set of $m \geq 2$ matrices of squared distances measured on the same group of n individuals, $\{\mathbf{\Delta}_\alpha\}_{\alpha=1,\ldots,m}$, and a vector of weights $\mathbf{w}$, the first requirement in the construction of the $\mathbf{w}$-joint metric is that all matrices $\mathbf{\Delta}_\alpha$ have the same *geometric variability*. This concept was introduced by Cuadras and Fortiana (1995) as a variant of Rao's diversity coefficient (Rao 1982a, Rao 1982b) and, given a squared distances matrix $\mathbf{\Delta}_\alpha = \left(\delta^2(\mathbf{z}_i, \mathbf{z}_j)\right)_{\{1 \leq i,j \leq n\}}$, its sample version in the weighted context is:

$$V_{\mathbf{w},\alpha}(\delta) = \frac{1}{2}\mathbf{w}'\,\mathbf{\Delta}_\alpha\,\mathbf{w} = tr(\mathbf{F}_{\mathbf{w},\alpha}), \tag{5}$$

where $\mathbf{F}_{\mathbf{w},\alpha}$ is the corresponding standardized inner-product matrix.
For each squared distances matrix $\{\mathbf{\Delta}_\alpha\}_{\alpha=1,\ldots,m}$, we consider its doubly $\mathbf{w}$-centered inner-product matrix $\mathbf{G}_{\mathbf{w},\alpha}$ and its standardized inner-product matrix

$$\mathbf{F}_{\mathbf{w},\alpha} = \mathbf{D}_{\mathbf{w}}^{1/2}\,\mathbf{G}_{\mathbf{w},\alpha}\,\mathbf{D}_{\mathbf{w}}^{1/2}, \quad \text{for } \alpha = 1,\ldots,m,$$

and obtain the $\mathbf{w}$-joint metric as that whose standardized inner-product matrix is:

$$\mathbf{F}_{\mathbf{w}} = \sum_{\alpha=1}^{m} \mathbf{F}_{\mathbf{w},\alpha} - \lambda \sum_{\alpha\neq\beta} \mathbf{F}_{\mathbf{w},\alpha}^{1/2}\,\mathbf{F}_{\mathbf{w},\beta}^{1/2}, \tag{6}$$

where λ is an interaction parameter that can be optimized according to several criteria. For example, in this work, we are interested in obtaining maximum explained variability in an Euclidean configuration, and in Section 4 we call optimum weighted related metric scaling to the Euclidean configuration obtained by this procedure. See Esteve (2003) for a wide and rigorous study on the construction of metrics.
The second summand of formula (6) is the key tool for eliminating redundant information coming from different sources (different variable types, in our case). Roughly speaking, this second term makes the difference with Gower's metric and provides the desired flexibility when dealing with mixed-type data. Formula (6) extends formula (8) of Cuadras (1998) to the weighted context, where the interaction parameter was fixed to $\lambda = 1/m$ and inner-product matrices were used instead of standardized inner-product matrices. Formula (6) is obtained so that the following properties are fulfilled when $\lambda = 1/m$. We explicit them for $m = 2$:

1. If $\mathbf{\Delta}_1 = \mathbf{0}$ then $\mathbf{F}_{\mathbf{w}} = \mathbf{F}_{\mathbf{w},2}$,
2. If $\mathbf{\Delta}_1 = \mathbf{\Delta}_2$ then $\mathbf{F}_{\mathbf{w}} = \mathbf{F}_{\mathbf{w},1} = \mathbf{F}_{\mathbf{w},2}$,
3. If the Euclidean configurations associated to $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ generate orthogonal subspaces on $\mathbb{R}^n$, then $\mathbf{F}_{\mathbf{w}} = \mathbf{F}_{\mathbf{w},1} + \mathbf{F}_{\mathbf{w},2}$,
4. $\mathbf{F}_{\mathbf{w}} \geq 0$.

Principal coordinates are computed directly from matrix $\mathbf{F}_{\mathbf{w}}$ of (6), but in case it is necessary, we can recover matrix $\mathbf{\Delta}$ with the following formula:

$$\mathbf{\Delta} = \mathbf{g}_{\mathbf{w}}\,\mathbf{1}' + \mathbf{1}\,\mathbf{g}_{\mathbf{w}}' - 2\,\mathbf{G}_{\mathbf{w}}, \tag{7}$$

where $\mathbf{G}_{\mathbf{w}} = \mathbf{D}_{\mathbf{w}}^{-1/2}\,\mathbf{F}_{\mathbf{w}}\,\mathbf{D}_{\mathbf{w}}^{-1/2}$ and $\mathbf{g}_{\mathbf{w}} = \mathit{diag}(\mathbf{G}_{\mathbf{w}})$.

4 A case study

In this Section we apply the two techniques described above to the data matrix $\mathbf{Z}$, whose rows are 84 records representing 13,296 dependent children with possibilities of receiving public aid (see Section 2.2 for the data set description). Hereafter, we call Gower's metric to the distance matrix derived using formula (4) and $\mathbf{w}$-joint metric to the distance matrix obtained from formula (7).
In this study, the $\mathbf{w}$-joint metric is constructed from $m = 3$ different squared distances matrices measured on the same set of $n = 84$ individuals. We call them $\mathbf{\Delta}_1$, $\mathbf{\Delta}_2$ and

$\mathbf{\Delta}_3$. The vector of weights, $\mathbf{w}$, is estimated by INE from the survey and taking into account the sampling design.
Matrix $\mathbf{\Delta}_1$ contains the information related to the three numerical variables considered in the study, that were the age of the child, the scale value reached by the child and weekly hours of attention. In particular, we compute $\mathbf{\Delta}_1$ matrix using a robust version of Mahalanobis' distance

$$\delta^2(\mathbf{z}_i, \mathbf{z}_j) = (\mathbf{z}_i - \mathbf{z}_j)' \, \mathbf{S}^{*-1} \, (\mathbf{z}_i - \mathbf{z}_j),$$

that consists of estimating the entries in the covariance matrix $\mathbf{S}^*$ in a robust way. The variance of the j-th continuous variable is estimated from a 5%-trimmed sample, as suggested by Tuckey (1960). A robust estimator for the covariance between variables Z_j and Z_k is obtained from

$$s^*_{jk} = \frac{1}{4}(\hat{\sigma}^{*2}_{+} - \hat{\sigma}^{*2}_{-}),$$

where $\hat{\sigma}^{*2}_{+}$ and $\hat{\sigma}^{*2}_{-}$ are robust estimators of the variances of $Z_j + Z_k$ and $Z_j - Z_k$, respectively (see Gnanadesikan 1997).
Matrix $\mathbf{\Delta}_2$ contains the information concerning three multi-state categorical variables, that is, information about the respondent, the type of received aids and the severity of limitations to perform ADL. In this case, we start by computing a similarity matrix $\mathcal{S}_2$, that contains Sokal-Michener's pairwise similarities, and obtain $\mathbf{\Delta}_2 = 2\,(\mathbf{1}\,\mathbf{1}' - \mathcal{S}_2)$.
Matrix $\mathbf{\Delta}_3$ contains the information of eighteen binary variables (sex and the limitations described in the Annex). In this case, we compute a similarity matrix $\mathcal{S}_3$, whose entries are the Jaccard's pairwise similarities, and obtain $\mathbf{\Delta}_3 = 2\,(\mathbf{1}\,\mathbf{1}' - \mathcal{S}_3)$.
Finally, from formula (6) we construct the $\mathbf{w}$-joint metric and obtain the Euclidean configurations shown in Figures 1–4. In this study, the interaction parameter λ is optimized so that these configurations have maximum explained variability.

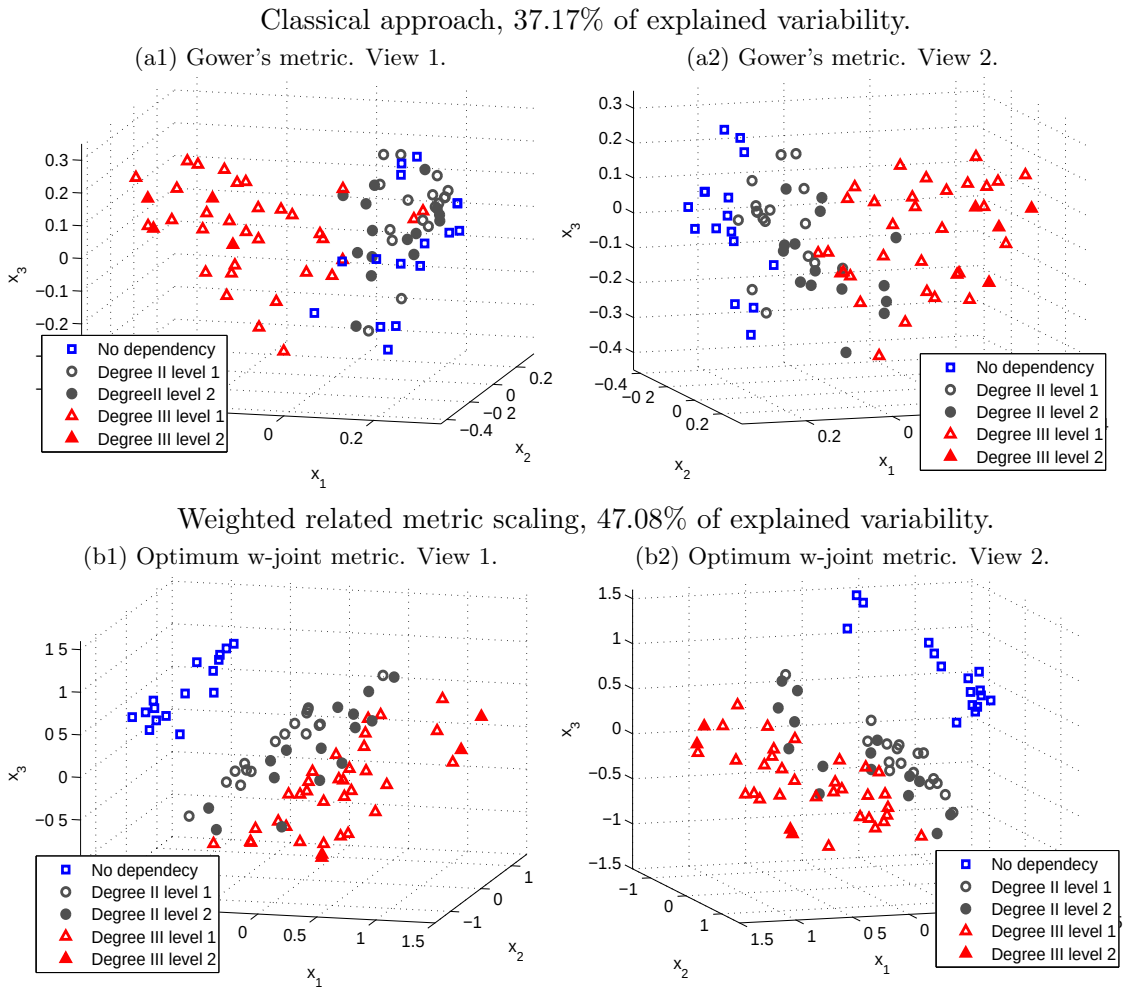
4.1 Euclidean configurations

In Figure 1 we depict three-dimensional principal coordinate representations of the data set obtained via the classical approach (panels (a1) and (a2)) and through weighted related metric scaling technique (panels (b1) and (b2)). We give two different views of each representation for better comparison.
Two main advantages can be observed while using the new proposal based on the $\mathbf{w}$-joint metric. Firstly, there is an increase in the percentage of explained variability and, secondly, the group of non-dependent children is quite well identified. A possible explanation may be that the information contained in the continuous variables is better incorporated with the $\mathbf{w}$-joint metric than with the classical approach. Hence, hereafter, and with the aim of defining homogeneous profiles of dependent children, we focus our attention in the $\mathbf{w}$-joint metric representation.

4.2 Looking for influent variables

Next, we are interested in capturing the underlying structure of the groups obtained in a *natural* way through weighted related metric scaling. To determine which variables are more powerful in explaining the homogeneity within groups we compute the

Figure 1: Euclidean maps obtained via (a) classical MDS and (b) Optimum weighted related metric scaling



the correlation coefficients between the original variables and the first three principal axes, shown in Table 5. We consider Pearson's correlation coefficient for continuous variables, whereas Spearman's correlation coefficient is computed for categorical variables.

Form Table 5 it can be seen that continuous variables, such as age, scale or hours-week, are more correlated with the principal axes obtained from the **w**-joint metric than with those obtained via Gower's metric. Moreover, categorical variable B5 and binary variables lim2, lim3, lim9 and lim11 also have influence on the principal axes. For instance, when using the **w**-joint metric, the first principal coordinate is mostly determined by variables lim 2, lim 3, lim 9, lim 11, B5, scale and hours-week, whereas age and hours-week are influent variables for the second principal axis. Finally, lim11, age and scale have great influence on the third principal coordinate. Such an information will be valuable in the definition of homogeneous profiles. Bearing this

Table 5: Correlation coefficients (Pearson for continuous variables and Spearman for categorical ones) between the principal coordinates and the considered variables.

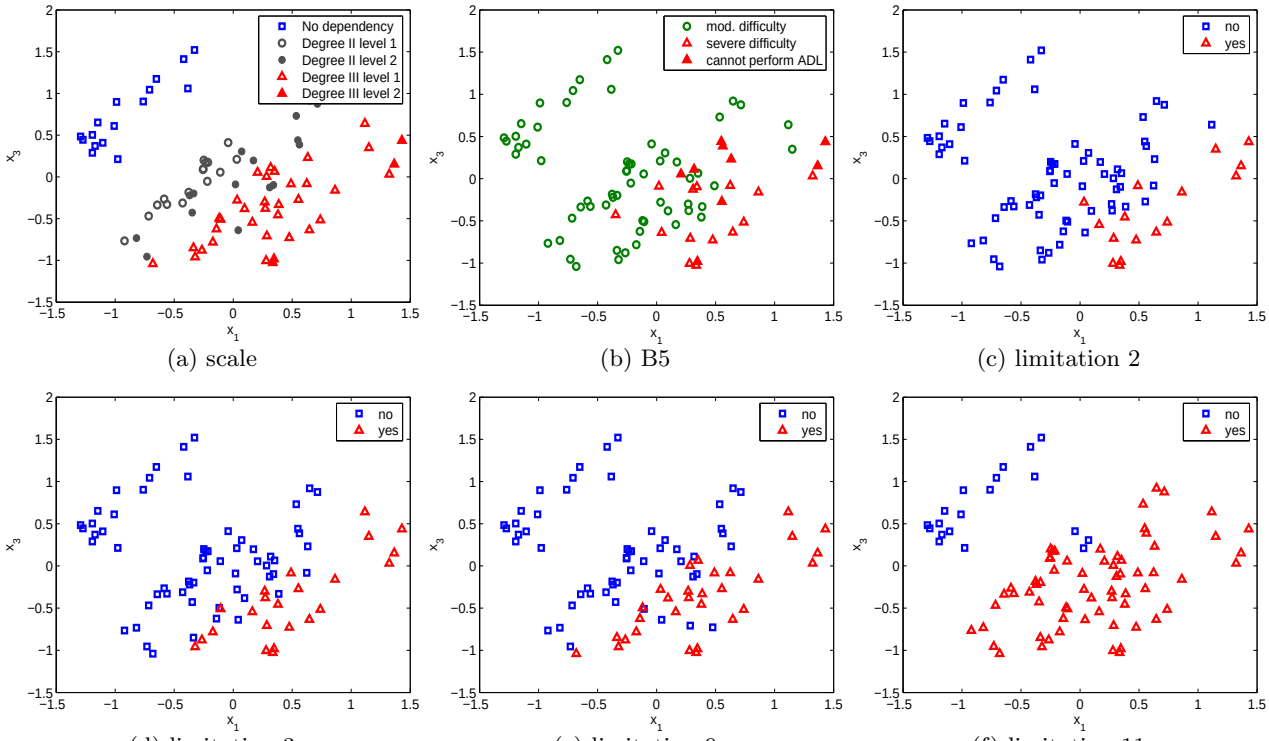
	Gower's metric			Optimum w-joint metric		
	1st P.C. (x_1)	2nd P.C. (x_2)	3rd P.C. (x_3)	1st P.C. (x_1)	2nd P.C. (x_2)	3rd P.C. (x_3)
sex	0.3187	0.3291	0.2791	-0.0844	0.1614	0.0771
lim 1	**-0.5261**	-0.0153	0.1754	0.3529	-0.2787	-0.3602
lim 2	**-0.6452**	0.0450	0.1500	**0.5226**	-0.2826	-0.2988
lim 3	**-0.7548**	-0.0622	0.0908	**0.5269**	-0.2890	-0.4046
lim 5	-0.3040	0.0550	0.1006	0.2376	-0.0197	-0.0322
lim 6	-0.0507	-0.0784	0.3159	-0.0438	-0.0922	-0.0715
lim 7	0.0409	0.0288	0.1031	-0.0667	0.0152	0.1167
lim 8	**-0.5914**	-0.2052	-0.0842	0.3816	-0.1556	-0.2870
lim 9	**-0.7910**	-0.1824	-0.0318	**0.5082**	-0.2490	-0.4591
lim 10	-0.3183	0.3550	-0.2097	0.3781	-0.0794	-0.0156
lim 11	-0.5475	0.4665	-0.0522	**0.5745**	0.1485	**-0.6378**
lim 12	-0.1840	0.1405	-0.4131	0.2392	-0.2743	0.0468
lim 13	-0.1332	**0.6302**	-0.4703	0.2756	0.0154	-0.0133
lim 14	-0.2475	**0.6663**	-0.1949	0.4333	-0.0048	0.0070
lim 15	-0.1164	0.2514	0.0755	0.0471	-0.2443	-0.3135
lim 16	0.0854	**0.7082**	0.4083	0.1135	0.2739	-0.0187
lim 17	0.1268	0.4842	**0.5499**	0.0427	0.2801	-0.0749
lim 18	-0.2546	0.3283	-0.2625	0.3271	0.0232	-0.0244
age	0.2712	0.0286	0.0448	0.2062	**0.6345**	**0.7274**
scale	**-0.6824**	**0.5046**	-0.0776	**0.7425**	0.1188	**-0.6761**
hours-week	**-0.5723**	0.1520	-0.4066	**0.6839**	**-0.5928**	0.0779
inf-relac	-0.1459	-0.1216	-0.1411	0.1557	0.0309	0.1409
B2	-0.3906	-0.3593	**0.5739**	0.1021	-0.1416	-0.2125
B5	-0.4657	0.4320	-0.1206	**0.5812**	-0.2005	-0.1699

Notes: Bold numbers reflect coefficients greater than 0.5 in absolute value.
Source: Own elaboration.

objective in mind, in Figures 2–3 we plot several projections (in dimension two) of the principal coordinate representations shown in panels (b1) and (b2) of Figure 1, using the information of those variables more correlated with the principal axes (lim 2, lim 3, lim 9, lim 11, B5, age, scale, hours-week) to color the individuals. In this way, groups of homogeneous individuals will become apparent.

After analysing all possible projections, we decided to include only the most representative ones. For this reason, Figure 2 contains the principal coordinate representation (3rd P.C. versus 1st P.C.) obtained from the **w**-joint metric. Looking at panel (f), and comparing it with panel (a), we can see that lim 11 (the child can hardly do the things that other children do at the same age) is crucial in splitting the individuals in two groups: dependants and non-dependants. In fact, when a child is declared as dependant, almost always lim 11 is present (95.1% and 90.8% in Degree II level 1 and 2, respectively, and 100% in Degree III). The remaining variables with correlation coefficients greater than 0.50 in absolute value (lim 2, lim 3, lim 9, age, scale, hours-week and B5) are quite useful for constructing dependency profiles. That is, children not affected by dependency show a moderate severity in limitations linked to ADL. Besides, they neither suffer those limitations associated to vertical movements (lim 2, lim 3 and lim 9) nor lim 11. Their scale value is fully identified using the 1st and 3rd principal corrdinates and, in most cases, their ages are between 48 and 71 months old.

Figure 2: Principal coordinate representations obtained via optimum weighted related metric scaling. Projections of Figure 1 (panels (b1) and (b2)) configurations onto 1st and 3rd P.C.



(a) scale (b) B5 (c) limitation 2

(d) limitation 3 (e) limitation 9 (f) limitation 11

The most important features in dependent children can be summarized as follows: none of them is under 50 points in the dependency scale; all exhibit moderate severity, at least; lim 2, lim 3 and lim 9 are not suffered by those with a scale value between 50 and 75 and they are hardly manifested in children with Degree III (37.9% and 5.2% in level 1 and 2, respectively). That is, these three limitations are associated to severe dependency cases. Summing up, we can establish the following profiles:

[Pr1] Non-dependent children (19%): Moderate difficulty to perform ADL, none of limitations 2, 3, 9, 11 are suffered, less than 56 weekly hours of attention are needed by 79% of them.

[Pr2] Severe dependent children (40.6%): Composed by 100% of children in Degree II (level 1,2). None of limitations 2, 3, 9 are suffered, almost all (93%) suffer limitation 11 and 59% of them need less than 56 weekly hours of attention.

[Pr3] Major dependent children (25.4%): Composed by 75% of children in Degree III level 1. All of them suffer limitations 9, 11 and 53% of them need more than 56 weekly hours of attention.

[Pr4] Utmost dependent children (15%): Composed by 100% of children in Degree III level 2 and 25% of children in Degree III level 1. Severe difficulties to perform

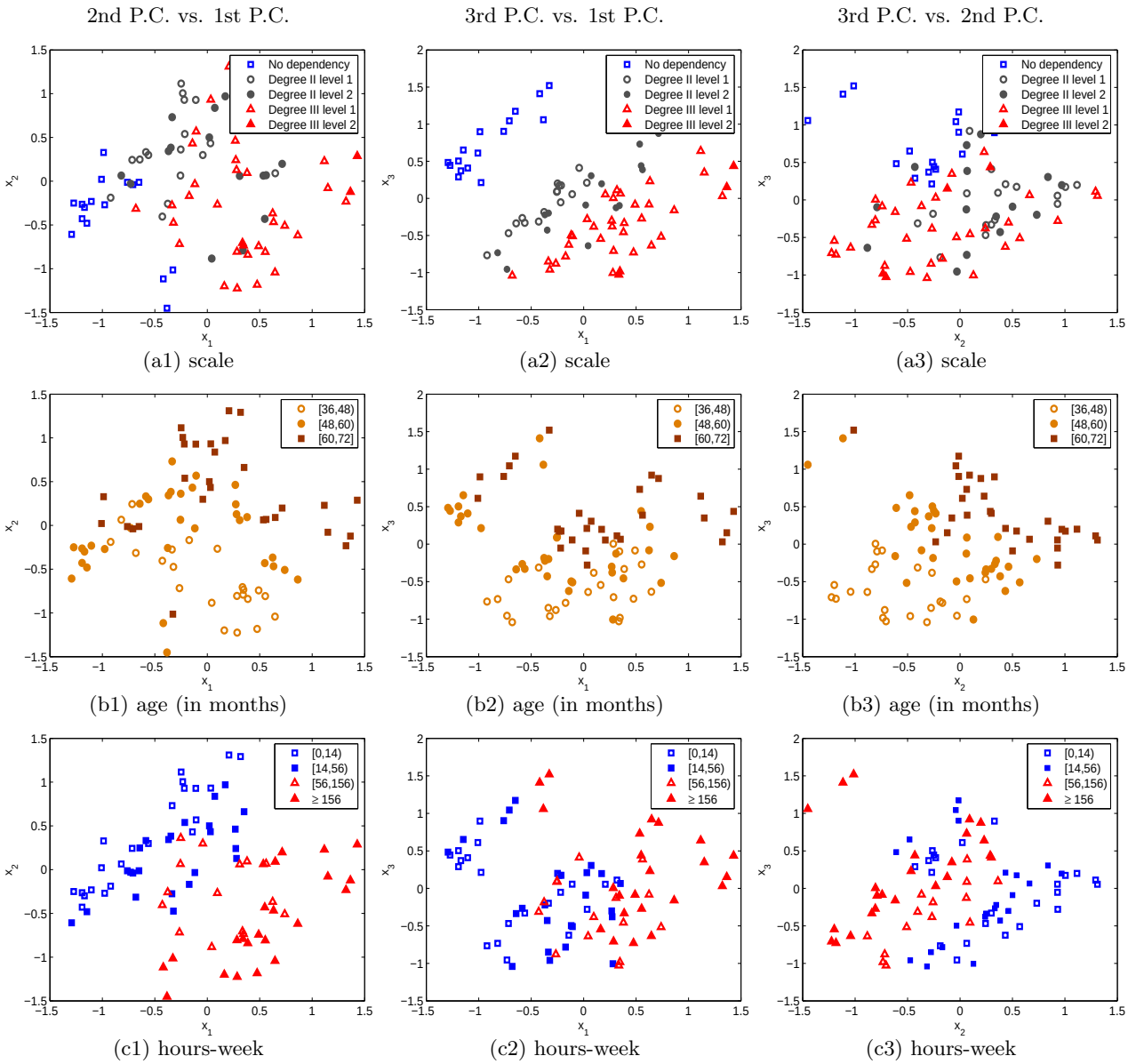
ADL or cannot perform them. All of them suffer limitation 11 and almost all (95%) suffer limitations 2, 3, 9, 82% of them need more than 56 weekly hours of attention, 88% of whom need more than 155 weekly hours of attention.

Profiles [**Pr1**] and [**Pr4**] clearly describe opposite situations, whereas profiles [**Pr2**] and [**Pr3**] consider different realities under the same Degree of dependency. In fact, it is possible to find individuals with the same scale value but with huge differences in difficulties to perform ADL, age and necessity of attention.
If we focus the attention on the relationship between the intensity of severities (B5) and the scale value reached by each child, we see that both variables are directly related. In fact, children in Degree II level 1 show a moderate severity and 78.8% of those in Degree III level 2 cannot perform ADL. On the other hand, looking at Figure 2 it seems that it is difficult to distinguish between severe difficulty to perform ADL and cannot perform ADL (there are groups of individuals with same values for limitations 2, 3, 9, 11, but not for B5). This may lead us to conclude that it would be better to join those categories in only one.
Figure 3 contains several projections of the principal coordinate representation obtained from the **w**-joint metric. Panels (a1)–(c1) depict 2nd P.C. versus 1st P.C., panels (a2)-(c2) show 3rd P.C. versus 1st P.C. and, finally, panels (a3)–(c3) contain 3rd P.C. versus 2nd P.C. We prefer to include again variable scale (panels (a1)–(a3)), for better comparison. For example, we can see the usefulness of variable age in panel (b3). A special case is that of variable hours-week, which seems to be contradictory with the groups defined by variable scale.
Despite the Spanish Act establishes a direct link between the amount of time devoted to care dependent people and the intensity of the dependency, one of the most surprising results is that there is a no direct relationship between the number of weekly hours for care and the level of dependency. In fact, there are individuals that need more than 155 hours per week in opposite situations (61% of children in Degree III level 2 versus 21.2% that are non dependent). See Figure 4 and also panels (a3) and (c3) of Figure 3. This same effect was noticed by Albarrán and Alonso (2006) and Gispert Magarolas, Clot-Razquin, Rivero Fernández, Freitas Ramírez, Ruíz-Ramos, Ruíz Luque, Busquets Bou, and Argimón Pallàs (2008). Similar results were found in Bihan and Martin (2006) when studying some European systems of assistance to dependent people. This fact reflects that the Spanish scale is not properly measuring the severity of the situation of dependency, which is quite worrying, since the scale value reached by an individual allows him/her to access to the benefits entitled by the Spanish Act 39/2006.

5 Concluding remarks

Disability is a large public health problem even in developed countries. Dependence is the main impact factor on health and quality of life. Suffering dependency when people is over 65 is the most common situation. For this reason, some authors join both states, dependency and ageing (see Casado and López 2001, Moragas and Cristofol 2003, and López(Dir.), Comas, Monteverde, Casado, Caso, and Ibem 2005). However, it is not true that all people over a certain age would be included in the group of population affected by this contingency. It would be possible to prevent

Figure 3: Principal coordinate representations obtained via optimum weighted related metric scaling. Projections of Figure 1 (panels (b1) and (b2)) configurations onto two principal axes.



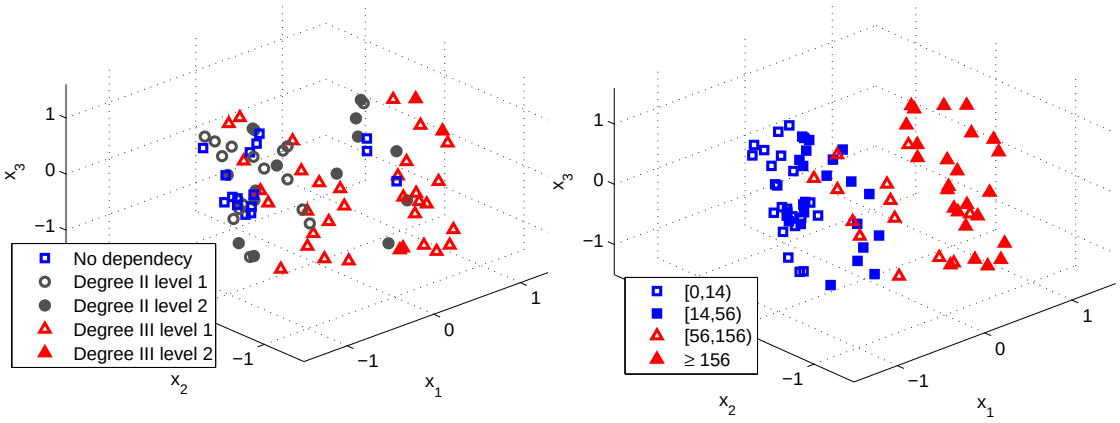
(a1) scale (a2) scale (a3) scale

(b1) age (in months) (b2) age (in months) (b3) age (in months)

(c1) hours-week (c2) hours-week (c3) hours-week

this situation if people followed a healthy way of life, the health care system was more efficient than today is and if we were able to have an early diagnosis of the chronical illness (Zunzunegui 1998). Although the former is true, it must be said that it is possible to find people in dependency at any age during the lifetime, even in the childhood. There are many studies about dependence when people is over 65, however, concerning children, there is a lack of research. This study contributes in this line. Our main purpose is to establish the existing relationship among different

Figure 4: Principal coordinate representations obtained via Optimum Weighted Related Metric Scaling. Individuals identified by (a) scale and (b) hours-week.



(a) scale

(b) hours-week

variables (numerical, categorical and binary) referred to Spanish children between 3 and 6 years old and their functional dependence in basic activities of daily living. Data comes from the Survey about Disabilities, Personal Autonomy and Dependence Situations, EDAD 2008, (Spanish National Institute of Statistics, 2008), where each individual represents a number of similar individuals. The number of multivariate techniques than can cope with mixed-type and weighted data is quite scarce. In this paper we propose a multivariate methodology for mixed-type data to search for homogeneous profiles. In particular, we extend the work of Grané and Romera (2009) to the weighted context. Moreover, we include an interaction parameter which provides more flexibility when dealing with mixed-type data. The main findings are: Firstly, this new technique overperfoms the classical one based on Gower's metric, in the sense that homogeneous groups are better separated. This may be due to the possibility of constructing an *ad-hoc metric* with the property of discarding redundant information. Secondly, the things that a child can hardly do compared with other children at the same age (limitation 11) seems to be crucial in splitting the individuals into dependants and non-dependants. This finding goes in the line of USDHHS (2004) and complements the universal definition and classification established by the International Classification of Functioning for children and youth (WHO 2001a). Thirdly, the time devoted to care dependent children and the intensity of the dependency is not directly related. This was also found by Albarrán and Alonso (2006), among others, and reinforces the finding that the Spanish scale is not properly measuring the severity of the situation of dependency. This is quite relevant, since the scale value reached by an individual allows him/her to access to the benefits entitled by the Spanish Act 39/2006.

References

Albarrán, I. and P. Alonso (2006). Dependent Individuals Classification based on the 1999 Disabilities, Impairments and Health Status Survey. *Revista Española de Salud Pública 80*(4), 349–360.

Albarrán, I., P. Alonso, and C. Bolancé (2009). A Comparison of the Spanish, the French and the German Valuation Scales to Measure Dependency and Public Support for People with Disabiities. *Revista Española de Salud Pública 83*(3), 379–392.

Anderberg, M. (1973). Cluster analysis for applications. Technical report, DTIC Document.

Benavente, G. and D. Pfeiffer (2002). Bibliography–an annotated bibliography on children with disabilities. *Disability Studies Quarterly 22*(2), 135–159.

Bihan, B. and C. Martin (2006). A comparative case study of care systems for frail elderly people: Germany, Spain, France, Italy, United Kingdom and Sweden. *Social Policy & Administration 40*(1), 26–46.

Boj, E., M. Claramunt, and J. Fortiana (2001). Herramientas estadísticas para el estudio de perfiles de riesgo. *Instituto de Actuarios Españoles, Tercera Época*(7), 59–89.

Boj, E., M. Claramunt, and J. Fortiana (2004). *Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación*, Volume 88. Cuadernos de la Fundación MAPFRE Estudios.

Borg, I. and P. J. F. Groenen (2005). *Modern Multidimensional Scaling: Theory and Applications* (second ed. ed.). New York: Springer.

Campbell, M. (1986). An integrated system for estimating the risk premium of individual car models in motor insurance. *ASTIN Bulletin 16*(2), 165–183.

Cannell, M., B. Brumback, E. Bouldin, J. Hess, D. Wood, P. Sloyer, J. Reiss, and E. Andresen (2011). Age group differences in healthcare access for people with disabilities: Are young adults at increased risk? results from the 2007 florida behavioral risk factor surveillance system. *Journal of Adolescent Health 49*, 219–221.

Casado, D. and G. López (2001). *Vejez, dependencia y cuidados de larga duración*. Barcelona: Fundación La Caixa.

Claeson, M. and R. J. Waldman (2000). The evolution of child health programmes in developing countries: from targeting diseases to targeting people. *Bulletin-World Health Organization 78*(10), 1234–1245.

Cox, T. F. and M. A. A. Cox (2000). *Multidimensional Scaling* (second ed. ed.). London, Boca Raton, FL.: Chapman & Hall.

Crews, J. and D. Lollar (2008). The public health dimensions of disability. *International Encyclopedia of Public Health*, 422–432.

Cuadras, C. M. (1998). Multidimensional dependencies in ordination and classification. In K. Fernández and A. Morineau (Eds.), *Analyses Multidimensionelles des Données*, Saint-Mand (France), pp. 15–25. CISIA-CERESTA.

Cuadras, C. M. and J. Fortiana (1995). A continuous metric scaling solution for a random variable. *Journal of Multivariate Analysis 52*, 1–14.

Cuadras, C. M. and J. Fortiana (1998). Visualizing categorical data with related metric scaling. In J. Blasius and M. Greenacre (Eds.), *Visualization of Categorical Data*, London. Academic Press.

Esteve, A. (2003). *Distancias estadísticas y relaciones de dependencia entre conjuntos de variables*. Ph. D. thesis, Universitat de Barcelona.

Giannakouris, K. (2009). Ageing characterises the demographic perspectives of the European societies. *Eurostat: Statistics in Focus. Retrieved 9*, 8–72.

Gispert Magarolas, R., G. Clot-Razquin, A. Rivero Fernández, A. Freitas Ramírez, M. Ruíz-Ramos, C. Ruíz Luque, E. Busquets Bou, and J. Argimón Pallàs (2008). Dependence Profile in Spain: An Analysis from the Disability Survey of 1999. *Revista Española de Salud Pública 82*(6), 653–665.

Gnanadesikan, R. (1997). *Methods for Statistical Data Analysis of Multivariate Observations*. New York: John Wiley & Sons.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics 27*, 857–874.

Gower, J. C. (1992). Generalized biplots. *Biometrika 79*, 475–493.

Gower, J. C. and D. Hand (1996). *Biplots*. London, UK: Chapman & Hall.

Grammenos, S. (2003). Feasibility Study. Comparable statistics in the area of care of dependent adults in the European Union. European Commission. Luxembourg.

Grané, A. and R. Romera (2009). Sensitivity and robustness in MDS configurations for mixed-type data: A study of the economic crisis impact on socially vulnerable Spanish people. Working Paper 10-35, Universidad Carlos III de Madrid. Statistics and Econometrics Series 19.

Grupo de Atención Temprana (2000). Real Patronato de Prevención y Atención a Personas con Minusvalía. Madrid.

Hauser-Cram, P., M. E. Warfield, and J. P. Shonkoff (2001). *Children with disabilities: A longitudinal study of child development and parent well-being*, Volume 66. Wiley-Blackwell.

Haveman, R. and B. Wolfe (2000). The economics of disability and disability policy. *Handbook of Health Economics 1*, 995–1051.

INE (2010). Encuesta sobre Discapacidad, Autonomía personal y Situaciones de Dependencia (EDAD), Metodología. Ed. Subdirección General de Estadísticas Sociales Sectoriales (INE), Madrid, España.

Jiménez, A. and A. Huete (2010). Estadísticas y otros registros sobre discapacidad en España. *Política y Sociedad 47*(1), 165–174.

Kamette, F. (2011). Dependency Care in the EU: a comparative analysis. Foundation Robert Schumann. Social Issues European Issue, 196.

Krzanowski, W. J. (1994). Ordination in the presence of of group structure for general multivariate data. *Journal of Classification 11*, 195–207.

Krzanowski, W. J. and F. H. C. Marriott (1994). *Multivariate Analysis. Part 1*, Volume Distributions, ordination and inference. London: Edward Arnold.

Lafortune, G. and G. Balestat (2007). Trends in severe Disability among elderly people: assessing the evidence in 12 OECD countries and the future implications. *Health Working Papers*.

Loimaranta, K., J. Jacobsson, and H. Lonka (1980). On the use of mixture models in clustering multivariate frequency data. In *Transactions of the 21st International Congress of Actuaries*, Volume 2, pp. 147–161.

López(Dir.), G., A. Comas, M. Monteverde, D. Casado, J. R. Caso, and P. Ibem (2005). *Envejecimiento y dependencia. Situación actual y retos de futuro*. Barcelona: Caixa Catalunya.

Millán-Calenti, J. (2006). *Principios de geriatría y gerontología*. McGraw-Hill Interamericana de España.

Moragas, R. and R. Cristofol (2003). *El coste de la dependencia al envejecer*. Barcelona: Herder.

Morgan, B. and A. Ray (1995). Non-uniqueness and inversions in cluster analysis. *Applied Statistics 44*, 117–134.

Pope, A. and A. Tarlov (1991). *Disability in America: Toward a national agenda for prevention*. National Academy Press, 2201 Consitution Ave., NW, Washington, DC.

Ramsay, J. O. (1980). Joint analysis of direct ratings, pairwise preferences and dissimilarities. *Psychometrika 45*, 149–165.

Rao, C. R. (1982a). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology 21*, 24–43.

Rao, C. R. (1982b). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhyā. The Indian Journal of Statistics, Series A 44*, 1–22.

Tuckey, J. W. (1960). A survey of sampling from contaminated distributions. In I. O. et al. (Ed.), *Contributions to Probability and Statistics*, pp. 448–485. Standford University Press.

USDHHS (2004). National agenda for children with special health care needs: Achieving the goals 2000. US Department of Health.

WHO (2001a). International Classification of Functioning, Disability and Health-Child and Youth version (ICF-CY). WHO (World Health Organization). Geneva.

WHO (2001b). International Classification of Functioning, Disability and Health (ICF). WHO (World Health Organization). Geneva.

Zunzunegui, M. (1998). *Envejecimiento y salud*. Madrid: Informe de la Sociedad Española de Salud Pública y Administración Sanitaria.

Annex I

Definition of limitations

lim 1	For children over 9 months old: the child has troubles to stay sitting down without help
lim 2	For children over 9 months old: the child has troubles to stay standing up without help
lim 3	For children over 9 months old: the child has troubles to walk by his/her own
lim 5	The child can hardly see
lim 6	The child is fully deaf
lim 7	It seems that the child can hardly hear
lim 8	The child has troubles to move his/her arms
lim 9	The child has any weakness or stiffness in the legs
lim 10	The child sometimes has convulsions, goes rigid or lose consciousness
lim 11	The child can hardly do the things that other children do at the same age
lim 12	The child is frequently sad or depressed
lim 13	The child can hardly mix with other children, as the children at the same age do
lim 14	For children over 2 years old: the child can hardly understand simple instructions
lim 15	For children between 2-3 years old: the child can hardly recognize and name objects
lim 16	For children between 3-5years old: the child can hardly speak
lim 17	The child is into any specialized education system for stimulation
lim 18	The child has been diagnosed by a doctor or a psychologist of any illness that last more than one year

Source: ICF-CY Classification