SYMBOLIC ANALYSIS OF THE TIME USE IN THE BASQUE COUNTRY

Marta Mas, Haritz Olaeta

EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

Donostia-San Sebastián, 1
01010 VITORIA-GASTEIZ
Tel.: 945 01 75 00
Fax.: 945 01 75 01
E-mail: eustat@eustat.es
www.eustat.es

SYMBOLIC ANALYSIS OF THE TIME USE IN THE BASQUE COUNTRY

Marta Mas, Haritz Olaeta

IATUR

Lisbon, 15-18 October 2002

Contents

Chapter
1

Introduction

Time use surveys provide very valuable information on what people do with their time, what proportion of time is used in economically productive activities, leisure pursuit, personal care or household related activities. Diversity in time use patterns between male and females, older and younger generations and any other different socio-economical groups will be important not only for policy development and programme planning but also for product and service production.

New statistical multivariate techniques to handle large amount of data are being developed and tested in the literature. As a member of the European project ASSO (e.g., Analysis System of Symbolic Objects, IST-2000-25161), EUSTAT (Institute of Statistics of the Basque Country), is currently testing a pilot software for symbolic data analysis (Sodas 1.2,[7]) that will be also presented and used in this work.

Symbolic Analysis techniques will be briefly introduced and applied to the Time Use Survey 1998 in the Basque Country in order to process, filter and identify subpopulations with significant differences in time use behaviour.

Chapter
2

Data description

Data refers to the last Time Use Survey performed in the Basque Country in 1998. Microfile consists of 5,040 individuals and 47 variables. These variables constitute a list of all activities carried out by individuals during the day. These activities are divided into various areas: physiological needs, work and studies, housework, family care, social life, active leisure and sports, passive leisure, and travelling.

Each week is divided into working days, Fridays, and weekends. This assures that all activities are included in the survey. Finally, data was collected in two steps, half of them in spring and the rest in autumn, in order to capture the seasonal nature of some activities.

Each individual is described by several socio-demographic and time use variables, the day of the week and the sampling weight.

	Socio-demographic var.				Time Use var.				
ID	Sex	Age	…	Day	Cleaning	Clothing	Sleep	...	Weight

Table 1 Data table of Time Use survey.

The set of socio-demographic variables used in this work is qualitative and can be described as follows:

- *Sex (2 mod: Male,* Female*)*
- *Age (3 mod: < 35 years old,* 35-*59 years old, > 59 years old)*

- *Relation to Labour Force (3 mod: Inactive, Employed, Unemployed)*

For time use variables there are two possibilities:

- Treat them as quantitative variables. Each variable represents time in minutes for each activity.
- Treat them as qualitative variables with 4 modalities. Each modality indicates the frequency in performing an activity:

    1. No participation
    2. Scarce participation
    3. Medium participation
    4. High participation

In order to synthesise and to make interpretations easier, in what follows the second approach is taken.

Chapter
3

Symbolic approach

A symbolic object is a group or a class of individuals that will be treated as a new statistical unit (for a general overview of Symbolic Analysis the reader is referred to [1]). These “new” second order units are constructed in this case by the Cartesian product of the modalities of the socio-demographic variables (i.e. sex, age and relation to Labour Force). Therefore, the maximum number of symbolic objects that can be built from the variables described above is 2*3*3 =18. Time use variables will describe each socio-demographic group by means of a probabilistic distribution computed taking into account the sampling weights.

The descriptions of two of the constructed Symbolic Objects are given below (in brackets the relative frequencies of each category):

```
OS "Male 35-59 years old Employed"(487) =
   [Sleep = {"Scarce Particip."(0.599551), "High
Particip."(0.102476), "Medium Particip."(0.297974)}]
  ^[Cooking = {"Scarce Particip."(0.189872), "Medium
Particip."(0.055235), "High Particip."(0.0113959), "No
Particip."(0.743497)}]
  ^[Cleaning = {"High Particip."(0.0191529), "Medium
Particip."(0.0179092), "No Particip."(0.921424), "Scarce
Particip."(0.0415141)}]
  ^[Clothing = {"Medium Particip."(0.00854806), "No
Particip."(0.956517), "Scarce Particip."(0.034935)}]
  ^[Shopping = {"Medium Particip."(0.0532089), "Scarce
Particip."(0.0921257), "High Particip."(0.0456937), "No
Particip."(0.808972)}]
  ^[Care_children = {"Scarce Particip."(0.0539605), "No
Particip."(0.902165), "Medium Particip."(0.0300215), "High
Particip."(0.0138527)}]
  ^[Care-elderly = {"Medium Particip."(0.00330496), "No
Particip."(0.991245), "High Particip."(0.00544966)}]
  ^[Read-Tv-Radio = {"Medium Particip."(0.313413), "Scarce
Particip."(0.403077), "No Particip."(0.108756), "High
Particip."(0.174754)}]
  ^[Personal_care = {"High Particip."(0.163538), "Scarce
Particip."(0.356883), "No Particip."(0.000966876), "Medium
Particip."(0.478612)}]
```

```
OS "Female 35-59 years old Employed"(182) =
   [Sleep = {"Scarce Particip."(0.708373), "High
Particip."(0.0552229), "Medium Particip."(0.236404)}]
  ^[Cooking = {"Scarce Particip."(0.355841), "Medium
Particip."(0.405077), "High Particip."(0.150505), "No
Particip."(0.0885766)}]
  ^[Cleaning = {"Medium Particip."(0.315934), "High
Particip."(0.144115), "No Particip."(0.262377), "Scarce
Particip."(0.277573)}]
  ^[Clothing = {"Scarce Particip."(0.241204), "High
Particip."(0.112188), "Medium Particip."(0.189236), "No
Particip."(0.457372)}]
  ^[Shopping = {"Medium Particip."(0.247616), "Scarce
Particip."(0.238755), "High Particip."(0.166888), "No
Particip."(0.346741)}]
  ^[Care_children = {"Scarce Particip."(0.1395), "No
Particip."(0.752471), "Medium Particip."(0.0871506), "High
Particip."(0.0208787)}]
  ^[Care-elderly = {"Medium Particip."(0.0191261), "No
Particip."(0.971512), "High Particip."(0.00936167)}]
  ^[Read-Tv-Radio = {"Medium Particip."(0.28024), "Scarce
Particip."(0.402758), "No Particip."(0.190325), "High
Particip."(0.126676)}]
  ^[Personal_care = {"Medium Particip."(0.38582), "High
Particip."(0.264413), "Scarce Particip."(0.349767)}]
```

These two groups are, *a priori*, supposed to have significant differences in housework participation. A graphical representation might help to identify them easily.

Chapter
4

Symbolic visualisation

The software for symbolic data analysis (Sodas 1.2, [7]) provides us a friendly graphical interface to represent symbolic objects. Each object is shown as a multiple-axis star containing the distribution of symbolic variables that describe the group (time use variables in our case).

Modalities with the highest frequency are joined by a line building up a modal polygon. Differences in shape and area between polygons will reflect substantial differences between objects.

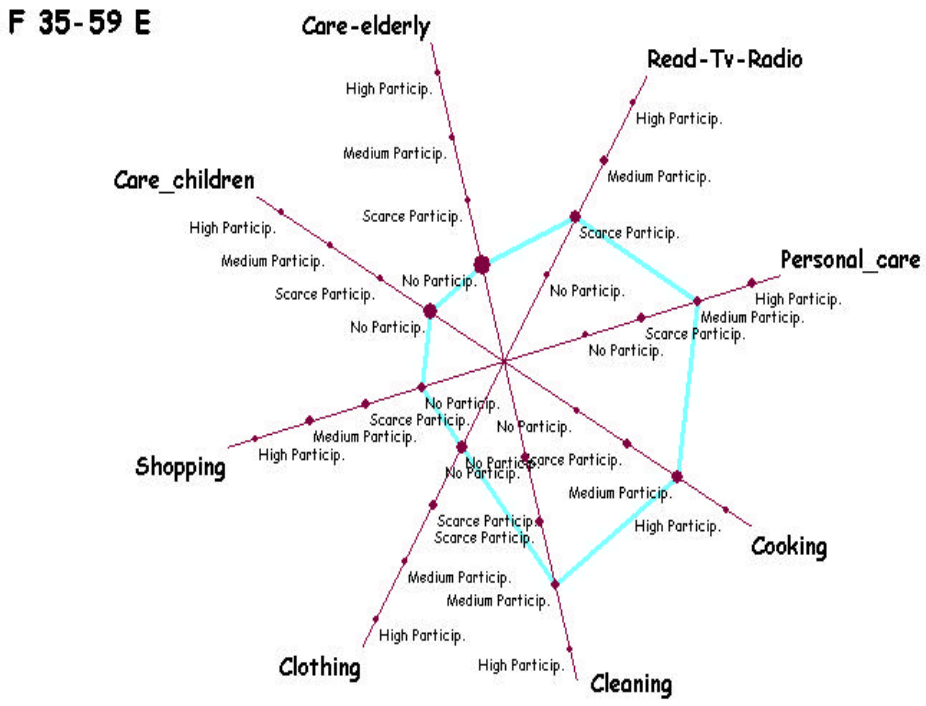

Figure 1. 2D- Zoom Star representation. Females. 35-59 years. Employed.

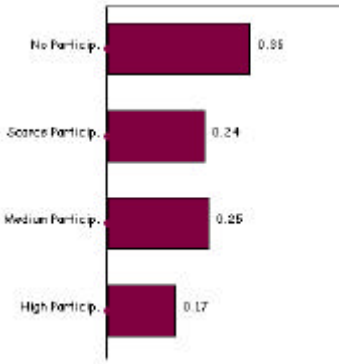

Figure 2. Univariate distribution for variable "Shopping".

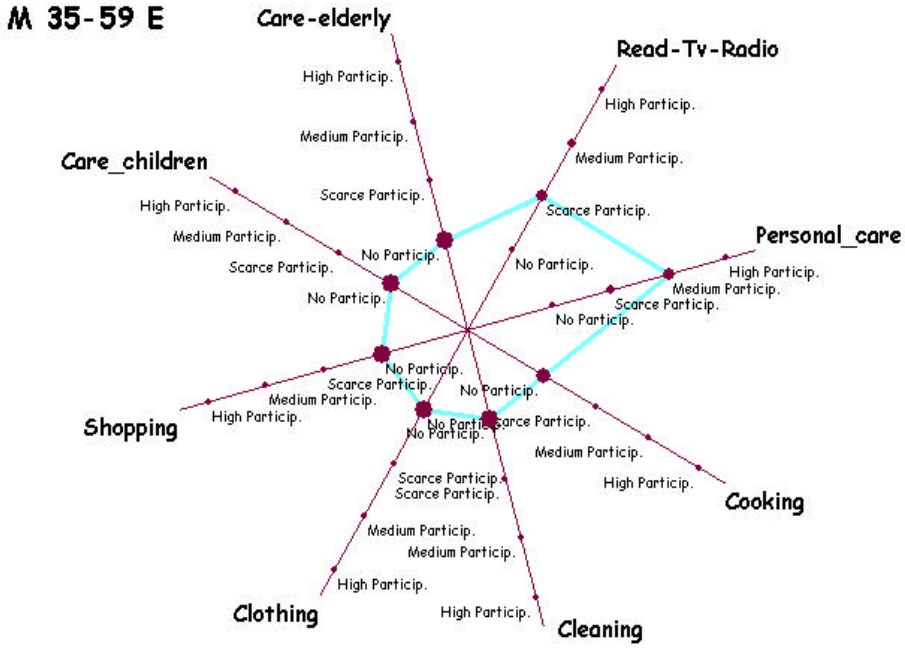

Figure 3. 2D- Zoom Star representation. Males.35-59 years. Employed.

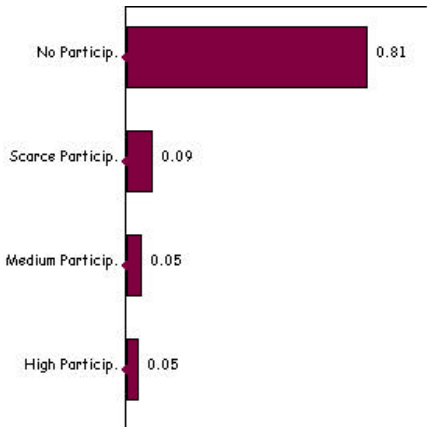

Figure 4.Univariate distribution for variable "Shopping".

The growing presence of women in the labour market has affected their participation in some of the housework and care activities. However, as it can be seen in Figures 1 and 3, women still spend much more time in housekeeping (cleaning, cooking,…) than men with the same socio-economical characteristics.

In addition, the highest concentration of men in "no participation" category, mainly in housework and care activities, contrast with a more heterogeneous distribution of women within all the modalities in variables like clothing or shopping. See Figures 2 and 4.

Chapter
5

Piramidal clustering

A clustering analysis has been done using PYR module [6] in SODAS 1.2. We have focused again in housework activities. Subpopulations with significant differences in housework behaviour will be identified and classified by a pyramid. The pyramidal clustering generalises hierarchies by allowing non-disjoint classes instead of a partition. Clusters are formed by considering variation on the values taken by the variables that describe the groups. In particular, symbolic objects are joined following the “Generality Degree” criterion (for details about this method see, for instance, [2]).

Three socio-demographic variables (i.e. sex; age and relation to Labour Force) have been used to construct eighteen symbolic objects defined by nine time use variables (i.e. sleep, personal care, read-TV-radio, care elderly, care of children, shopping, clothing, cleaning, cooking). For this analysis only data referring to working days have been considered.

The symbolic objects described above constitute the input data for the method and, therefore, a *symbolic* pyramid is constructed. However, a *numerical* (classical) pyramid using previously computed distance or dissimilarity matrices, could have been used as an input.

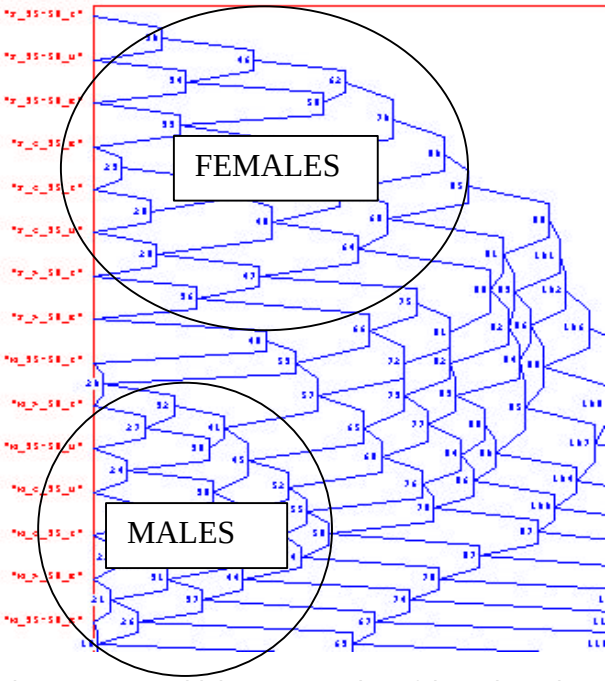

Figure 5.- Pyramidal representation of the selected symbolic objects.

The graphical output clearly shows two gender clusters and also a natural order, given by the method, which determines the "closest" symbolic objects. Employed men between 16 and 59 years old are joined in a first step (cluster 19), while the first "female cluster" (cluster 23) consists of women under 35 either employed or inactive. Age groups in women affect more to housework activity behaviour than the economical situation.

In general, male objects are "closer" than female ones (see differences in pyramid volumes in Figure 5). This reveals a quite homogeneous behaviour of men in relation to housework activities but a changeable situation in the case of women.

Each node in the pyramid has its own symbolic description. See below the description of an intermediate node that contains the closest male-female objects:

```
Node 48 ={"Males_35-59 years old_Inactive",
    "Females_>_59 years old_Employed"}

Node 48=[Sleep=(1(0.3191),2(0.2669),3(0.5326))]^
[Cooking=(1(0.6392),2(0.5343),3(0.1225),4(0.1942))]^
[Cleaning=(1(0.8099),2(0.2730),3(0.1278),4(0.2176))]^
[Clothing=(1(0.9844),2(0.3240),3(0.0797),4(0.0299))]^
[Shopping=(1(0.6051),2(0.2036),3(0.3698),4(0.1971))]^
[Care_children=(1(0.9911),2(0.0000),3(0.0155),4(0.0089))]^
[Careelderly=(1(0.9963),2(0.0070),3(0.0000),4(0.0037))]^
[Read-Tv Radio=(1(0.5324),2(0.3828),3(0.5134),4(0.0707))]^
[Personal_care=(1(0.5145),2(0.3022),3(0.4095),4(0.0461))]
```

Notice that frequencies in symbolic variables are not a probabilistic distribution any more but an accumulative distribution. Values in the modalities reflect the “at most” proportion of the population that presents this modality.

Chapter
6

Conclusions

We have shown an approach that is able to deal with large amount of data and provides a software (Sodas 2.1) that is at the same time easy to use and friendly to interpret. A descriptive analysis and a clustering division have been performed to point again the gender differences in time use, especially in housework activities.

Chapter 7

References

[1] BOCK, H.H., DIDAY, E., ET AL.

Analysis of Symbolic Data. Springer-Verlag. (2000)

[2] BRITO, P.

Symbolic Data Analysis: Inheritance Pyramids. University of Paris IX Dauphine (1991).

[3] CALVO, P.

Applications of Symbolic Objects in Official Statistics. Technical Report. EUSTAT (2000).

[4] . NOIRHOMME-FRAITURE, M. AND ROUARD, M.

Zoom Star: a solution to complex statistical object representation. In St. Howrd, J. Hammond and G. Lindgaard (eds): Proc. INTERACT 97, Sidney. (1997 a)

[5] NOIRHOMME-FRAITURE, M. AND ROUARD, M.,.

Computer graphics for symbolic objects. Presented at the Conference on Applied Stochastic Models and Data Analysis (ASMDA 97), Napoli, Italy (1997 b).

[6] OLDEMAR, R.

PYR: Pyramid *Clustering*

LISE-CEREMADE

University Paris IX Dauphine

[7] SODAS project

SODAS 2.1 Reference Manual.

http://www.ceremade.dauphine.fr/~touati/sodas-pagegarde.htm