

APPROACHES TO IMPUTATION OF INCOME FROM THE PERSPECTIVE OF CLASS ANALYSIS. A PROPOSAL USING FACTOR ANALYSIS TECHNIQUES.

RC55 Social Indicators
Imputation and Social Indicators: The Use of Factor Analysis for Imputing Missing Data



María Clara Fernández Melián.

Becaria doctoral del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET).

Instituto de Investigaciones Gino Germani (IIGG)
Universidad de Buenos Aires

José Javier Rodríguez de la Fuente.

Becario doctoral del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET).

Instituto de Investigaciones Gino Germani (IIGG)
Universidad de Buenos Aires



OBJECTIVE

The purpose of the paper is to analyze the implications of using different methods to input missing income data from the perspective of structure and class analysis. In particular, the interest is to describe and apply different methods and to examine the similarities and differences between them. In this way, it is expected to contribute to the problem of the lack of response in questions about income in household surveys .

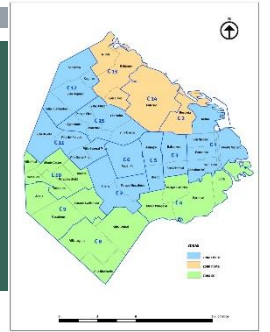
Methods:

- Assignment of median income value according to social class.
- Multiple correspondence analysis and assignment of median income value according to clusters.

METHODOLOGY

- Survey: “Social mobility and opinions about current society” 2012-2013. FONCYT project “Tendences and transformations in social structure: the impact of processes of social mobility in consumption horizons and political participation. An analysis of Metropolitan Region of Buenos Aires” coordinated by Dr. Eduardo Chávez Molina (Gino Germani Research Institution- University of Buenos Aires).
- Universe of analysis: people aged 30 years and above, occupied and residing in the City of Buenos Aires.

METHODOLOGY



Active variables:

- **Social class** according to Susana Torrado (1992, 1998) classification (occupation, occupational category and size of establishment).
- **Consum** considering possession of: TV LCD/LED, notebook, game console, Refrigerator with freezer, cellphone, internet connection, satellite TV, Blu-Ray, mattresses, air conditioner and stove with oven.
- **Residence zone** according to the classification of the General Direction of Statistics and Census of CABA. North: Commune 2, 13 and 14; Center: Commune 1, 3, 5, 6, 7, 11, 12 and 15 and South: Commune 4, 8, 9 and 10. Nivel educativo.

Supplementary variable:

- **Income:** 1) up to \$1.400 (95 USD), 2) \$1.401 to \$3.000 (202 USD), 3) \$3.001 to \$6.000 (405 USD), 4) \$6.001 to \$12.000 (811 USD), 5) \$12.001 to \$24.000 (1.622 USD) and 6) more than \$24.001.

METHODOLOGY

Table N° I. Complete and aggregate version of social class squeme.

Complete	Aggregate
I. Company directors	Company directors (I)
II. Proffesionals in specific function autonomous	Proffesionals in specific function (II y III)
III. Proffesionals in specific function employee	
IV. Owners of small companies	Owners of small companies and Small producers autonomous (IV y V)
V. Small producers autonomous	
VI. Technicians and assimilated groups	Technicians and assimilated groups(VI)
VII. Administratuve employees and sellers	Administratuve employees and sellers (VII)
VIII. Specializes autonomous employees	Skilled workers(VIII y IX)
IX. Skilled workers	
X. Non-skilled workers	Non-skilled workers(X, XI y XII)
XI. Non-skilled autonomous	
XII. Domestic employees	
Without specification	Without specification

Source: own elaboration based on Torrado scheme (1994, 1998), Boado (2008) and Pla (2012)

CHARACTERIZATION OF MISSING VALUES

Table N° 2. Cases with and without missing data according to age

Age	Missing data	No missing data
30-39	55 (32)	195 (37)
40-49	51 (29)	143 (27)
50-59	47 (27)	133 (25)
60-69	14 (8)	45 (9)
70-79	7 (4)	9 (2)
80-89	0 (0)	1 (0)
Total	174 (100)	526 (100)

Table N° 3. Cases with and without missing data according to sex

Sex	Missing data	No missing data
Varón	87 (50)	258 (49)
Mujer	87 (50)	268 (51)
Total	174 (100)	526 (100)

Note: parenthesis included percentages values.

Source: own elaboration based on Survey FONCYT 2012-2013

CHARACTERIZATION OF MISSING VALUES

Table N° 4. Cases with and without missing data according to level of education

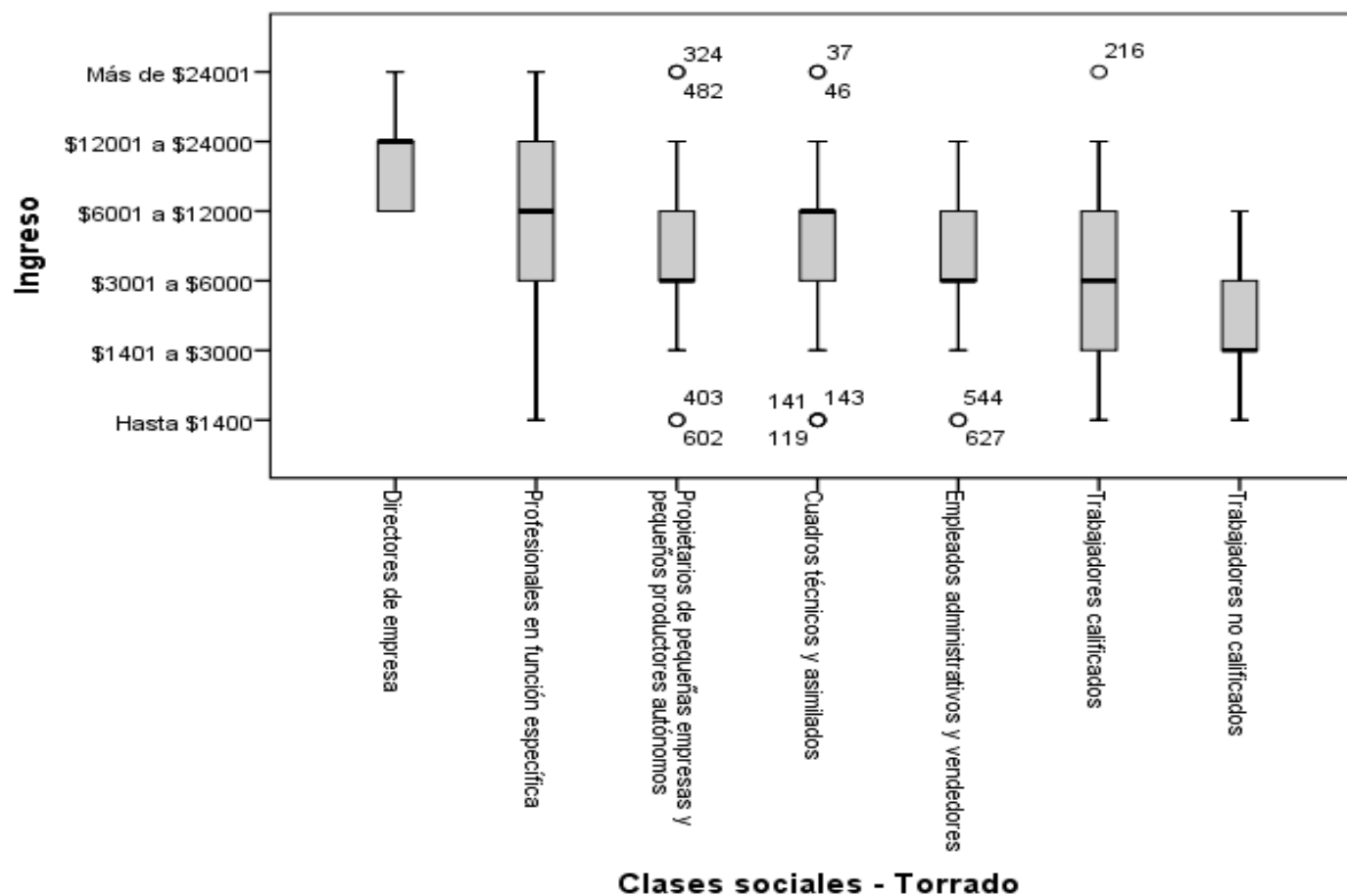
Education	Missing data	No missing data
Complete primary	38 (22)	132 (25)
Complete secondary	74 (43)	212 (40)
Complete higher education or more	62 (36)	182 (35)
Total	174 (100)	526 (100)

Table N° 5. Cases with and without missing data according to occupational category

Occupational category	Missing data	No missing data
Employer	10 (6)	44 (8)
Self-employed	57 (33)	171 (33)
Employee	96 (55)	286 (54)
Worker without salary or family assistance	2 (1)	4 (1)
Domestic service	8 (5)	21 (4)
Total	174 (100)	526 (100) ⁷

INCOME AND SOCIAL CLASS

Graph N° I. Box plot of income according to social class



Note: up \$1.400 (95 USD), \$1.401 to \$3.000 (202 USD), \$3.001 to \$6.000 (405 USD), \$6.001 to \$12.000 (811 USD), \$12.001 to \$24.000 (1.622 USD) and more of \$24.001 (1.622 USD).

Source: own elaboration based on Survey FONCYT 2012-2013

INCOME AND SOCIAL CLASS

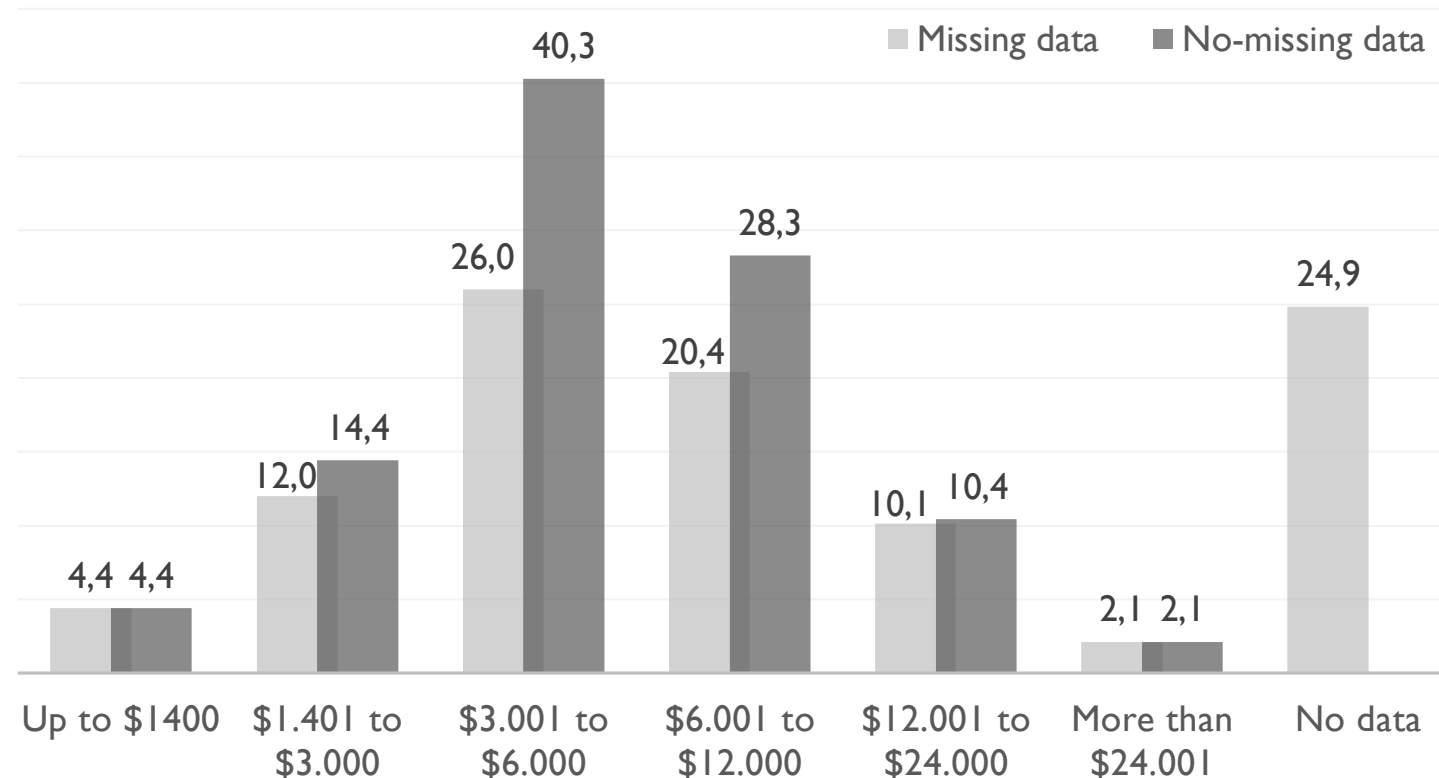
Table N° 6. Income medians according to social class

Social class	Median
Company directors	5 (\$12.001 to \$24.000)
Professionals in specific function	4 (\$6.001 to \$12.000)
Owners of small companies and Small producers autonomous	3 (\$3.001 to \$6.000)
Technicians and assimilated groups	4 (\$6.001 to \$12.000)
Administrative employees and sellers	3 (\$3.001 to \$6.000)
Skilled workers	3 (\$3.001 to \$6.000)
Non-skilled workers	2 (\$1.401 to \$3.000)
Total	3 (\$3.001 to \$6.000)

Source: own elaboration based on Survey FONCYT 2012-2013

IMPUTATION OF MISSING INCOME DATA ACCORDING TO SOCIAL CLASS

Graph N° 2. Percentage of income according to missing-data



MULTIPLE CORRESPONDENCE ANALYSIS

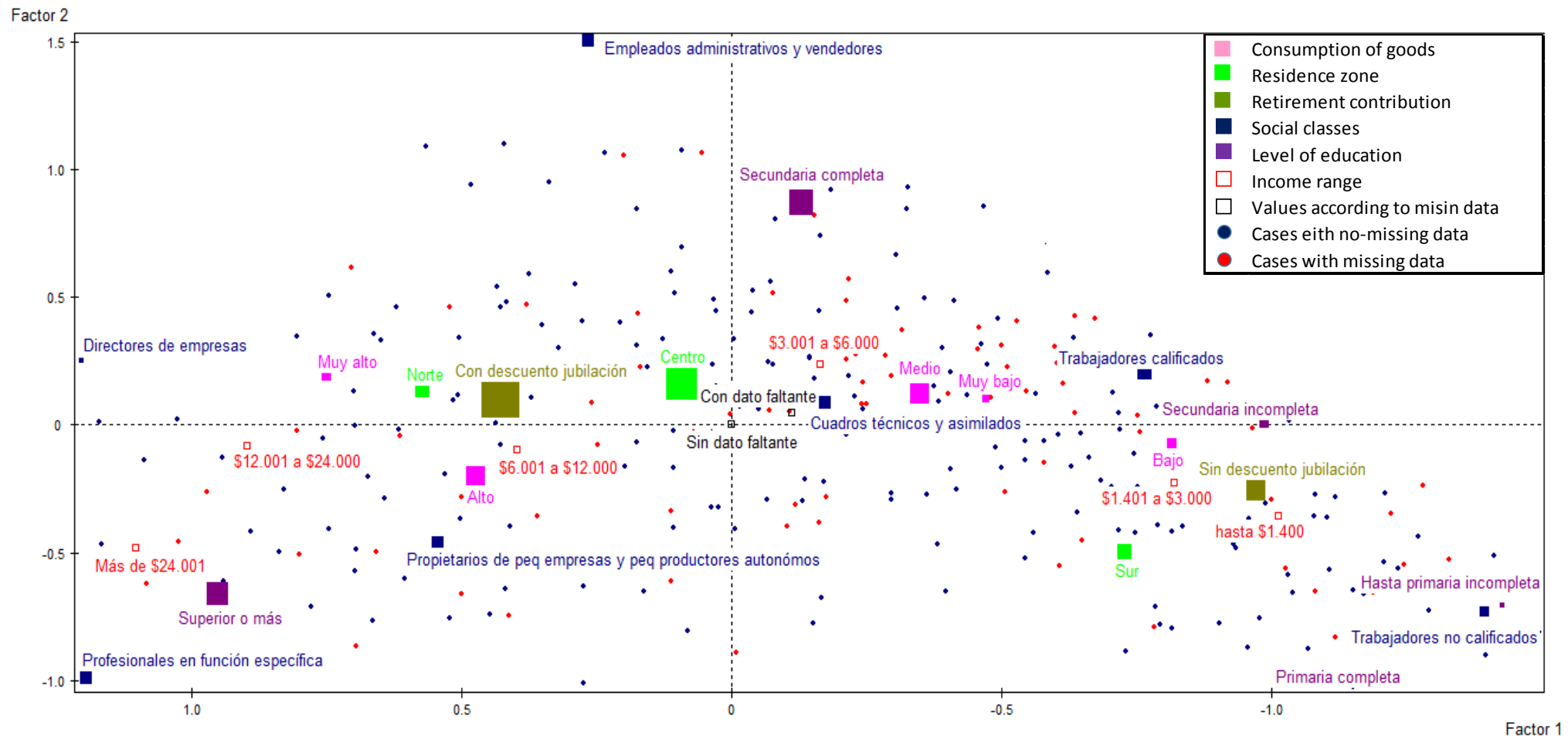
Table N° 7. Selected variables: own corrected values and percentage of explained inertia.*

Factor	Eigenvalue	% Inertia	Corrected own value(*)	% Inertia (1)	% Inertia (2)	Acumulated %
1	0,449600	12,49%	0,097344	88,70%	73,64%	73,64%
2	0,266537	7,40%	0,006917	6,30%	5,23%	78,87%
3	0,234561	6,52%	0,001866	1,70%	1,41%	80,28%
4	0,231797	6,44%	0,001580	1,44%	1,20%	81,48%

Note: * This percentage was calculated using Greenacre correction (2008). Using Benzécri's correction (1979), the explained variance with two first factors is 89%.

MULTIPLE CORRESPONDENCE ANALYSIS

Graph N° 3. Factorial analysis: active and supplementary variables



ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

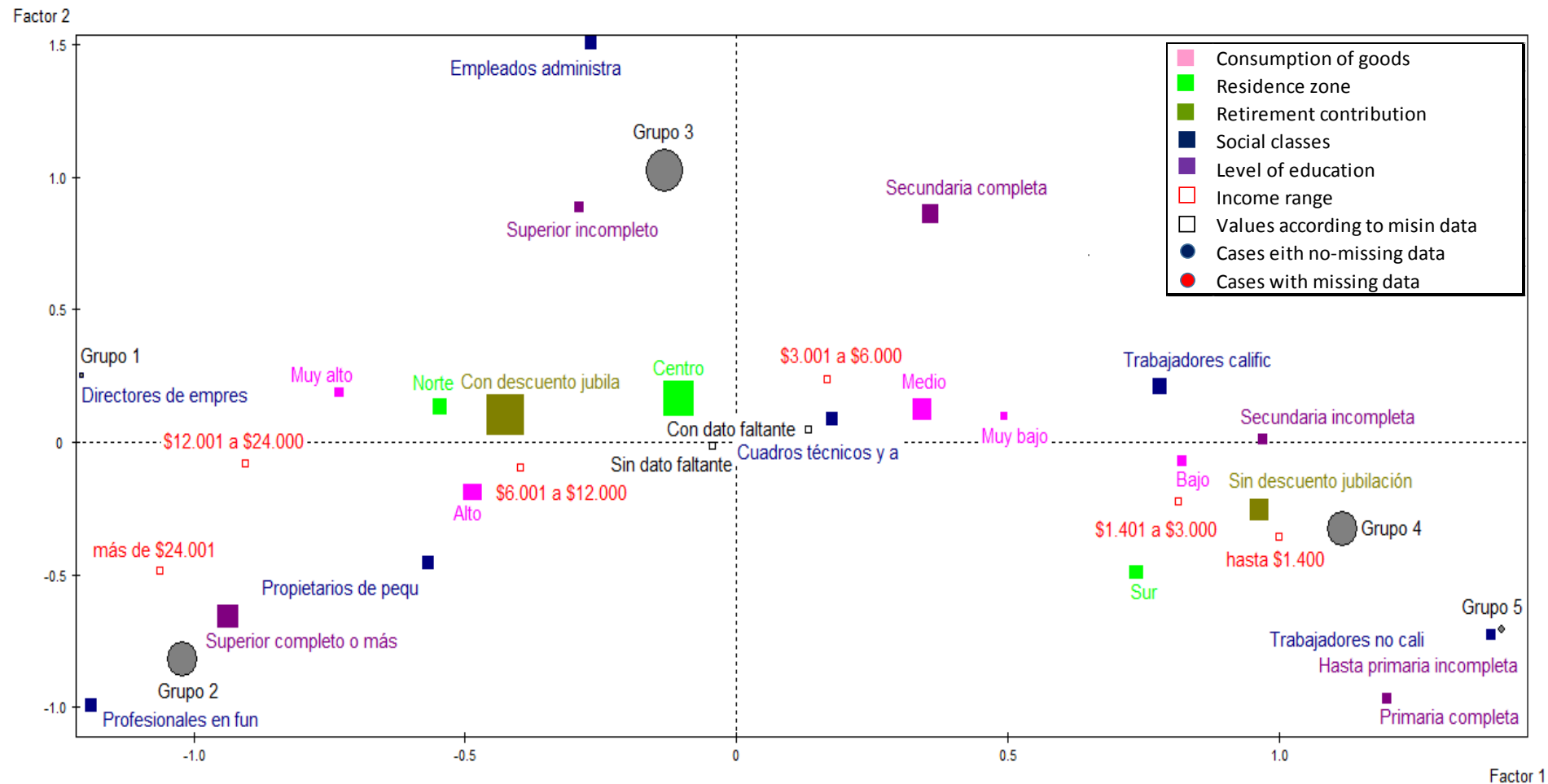
The **first factorial axis** explains 73.6 % of inertia and distributed to the population based on differences in **social class, area of residence , education level and perception of retirement contributions** (which can be thought of as an indicator labor formality).

Active variables: at one end (negative scores) there are those workers who are unskilled, without retirement contribution, with complete primary and living in southern part of the CABA, while at the other end there are those who are professions in specific function and to a lesser extent managers with higher education level or more, living in northern and with retirement savings.

Illustrative variables: the axis located in an end (negative) to those who have low-income, under \$ 3,000 and at the other end to those who receive incomes of \$ 12,000 or more. Note that the variable on consumption of goods was not significant for the analysis of the sample according differentiation factor analysis.

MULTIPLE CORRESPONDENCE ANALYSIS

Graph N° 4. Factorial analysis and classification



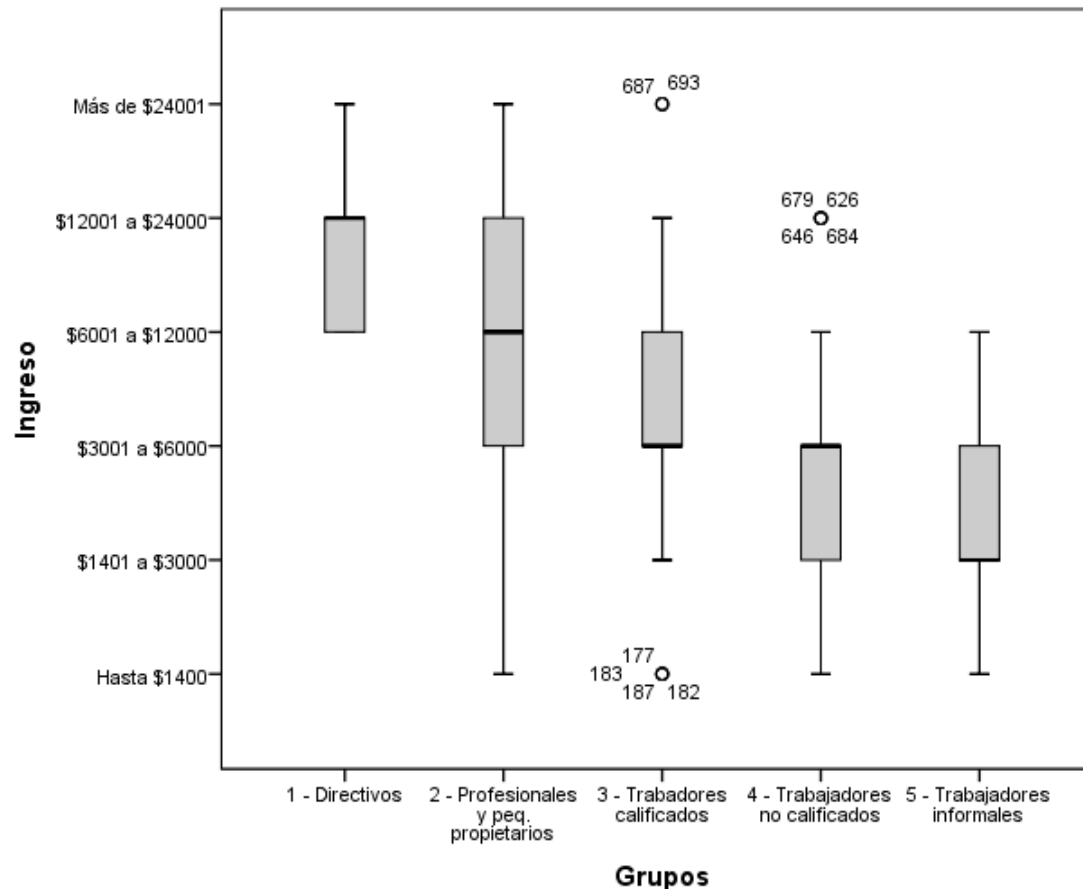
MULTIPLE CORRESPONDENCE ANALYSIS

Clusters:

- **“Managers”**: company directors residing in north zone.
- **“Professionals and small business owners”**: professionals in specific function and small business owners, with retirement contribution and higher education level.
- **“Skilled workers”**: administrative workers, salesmen and technicians, with complete secondary school, with retirement contribution and living in the north zone.
- **“Non-skilled workers”**: non-skilled workers, with incomplete primary and residing in southern zone.
- **“Informal employees”**: workers without retirement contribution, with complete primary and who live in south zone.

CLUSTERS AND INCOME

Graph N° 3. Box plot of income according to clusters.



Note: up \$1.400 (95 USD), \$1.401 to \$3.000 (202 USD), \$3.001 to \$6.000 (405 USD), \$6.001 to \$12.000 (811 USD), \$12.001 to \$24.000 (1.622 USD) and more of \$24.001 (1.622 USD).

CLUSTERS AND INCOME

Table N° 7. Income medians according to clusters

Clusters	Median
1 – Managers	5 (\$12.001 to \$24.000)
2 - Professionals and small bussiness owners	4 (\$6.001 to \$12.000)
3 - Skilled workers	3 (\$3.001 to \$6.000)
4 - Non-skilled workers	2 (\$1.401 to \$3.000)
5 - Informal employees	3 (\$3.001 to \$6.000)

Source: own elaboration based on Survey FONCYT 2012-2013

IMPUTATION OF MISSING DATA ACCORDING TO MULTIPLE CORRESPONDENCE ANALYSIS

Graph N° 4. Percentage of income according to missing data



COMPARISON OF RESULTS OF BOTH METHODS

Table N° 8. Percentage according to income and selected method of imputation

Income	Imputation method	
	Social class	ACM
Up to \$1400	4,4	4,4
\$1.401 to \$3.000	14,4	12,4
\$3.001 to \$6.000	40,3	43,1
\$6.001 to \$12.000	28,3	27,4
\$12.001 to \$24.000	10,4	10,4
More than \$24.001	2,1	2,1

Source: own elaboration based on Survey FONCYT 2012-2013

FINAL COMMENTS AND CONSTRAINTS

- **Imputation according to social class results in a greater size in the categories of low income** in relation to the clusters imputation (which implies an increase in the average category), although the extreme categories are held constant.
- **The use of the ACM allows to realize about the relevance of other variables** for the study of social differentiation such as retirement contribution with its consequent link with the labor formality (very relevant issue in Argentina and Latin America), in addition to the educational level and the area of residence.
- It should be noted that while the schemes of social class does not directly reflect differences in income, **occupation variable have a central place in the analysis of social stratification** as defined groupings with a culture and a set of common interests. And in addition, other variables considered relevant for the analysis of inequality such as education and income are related to a greater or lesser extent with that. (Francés García, 2009).
- **The income variable in range is a limitation** that is expected to solve in further analyzes with other sources of information for comparative studies. However, it was not possible to reconstruct it (as numeric) due to response problems.