

# Approaches to imputation of income from the perspective of class analysis. Aproposal using factor analysis techniques.

Fernández Melián, María Clara y Rodríguez de la Fuente, José.

Cita:

Fernández Melián, María Clara y Rodríguez de la Fuente, José (Julio, 2016). *Approaches to imputation of income from the perspective of class analysis. Aproposal using factor analysis techniques. 3rd ISA Forum "The Futures We Want: Global Sociology and the Struggles for a Better World". International Sociological Association, Viena.*

Dirección estable: <https://www.aacademica.org/maria.fernandez.melian/10>

ARK: <https://n2t.net/ark:/13683/pQ2q/ckP>



Esta obra está bajo una licencia de Creative Commons.  
Para ver una copia de esta licencia, visite  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

*Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.*

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



**María Clara Fernández Melián. Licenciada en Sociología. Becaria doctoral del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET). Instituto de Investigaciones Gino Germani (IIGG), Universidad de Buenos Aires (UBA). Docente en Universidad Nacional de Tres de Febrero (UNTREF)**

**José Javier Rodríguez de la Fuente. Licenciado en Sociología. Becario doctoral CONICET, IIGG, UBA.**



## 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 .

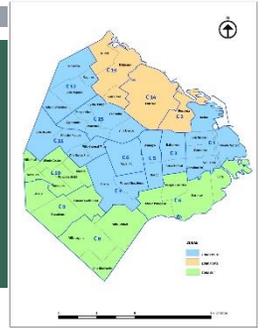
### Methods:

- 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, mattress, 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.**

<b>Complete</b>	<b>Aggregate</b>
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. Administrative employees and sellers	Administrative 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**

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

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

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

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**

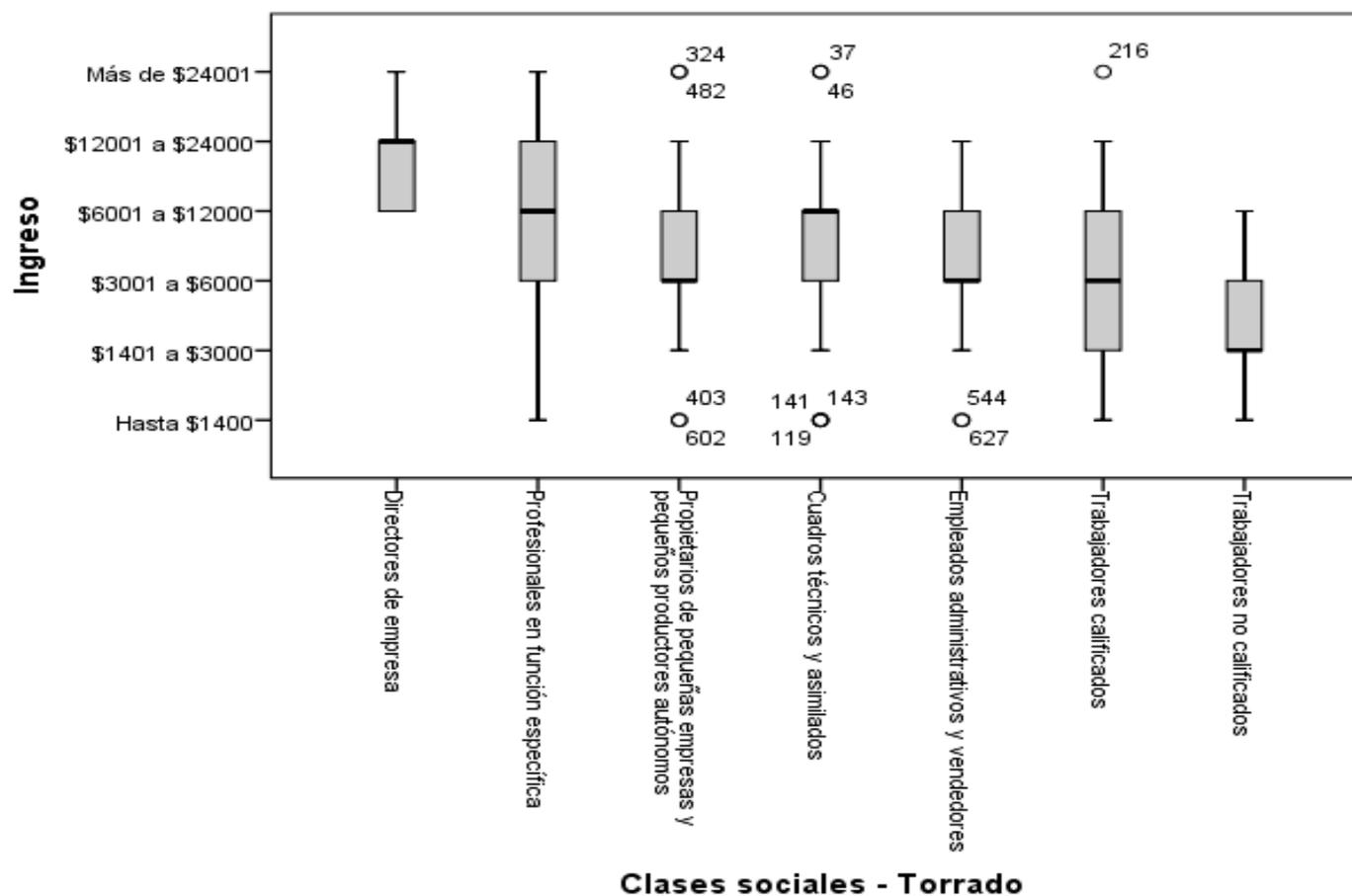
Education	Missing data	No missing data
<b>Complete primary</b>	38 (22)	132 (25)
<b>Complete secondary</b>	74 (43)	212 (40)
<b>Complete higher education or more</b>	62 (36)	182 (35)
<b>Total</b>	174 (100)	526 (100)

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

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

# 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

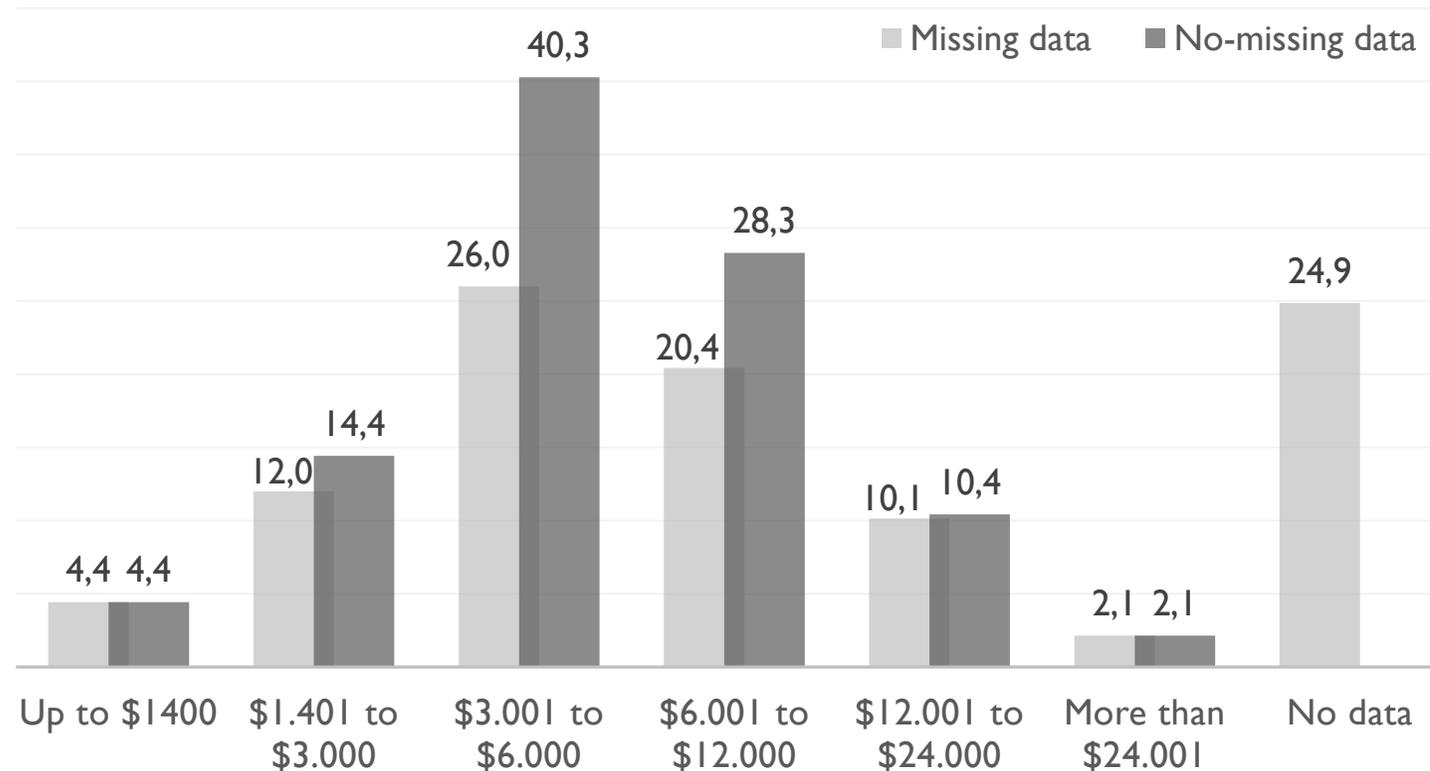
**Table N° 6. Income medians according to social class**

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

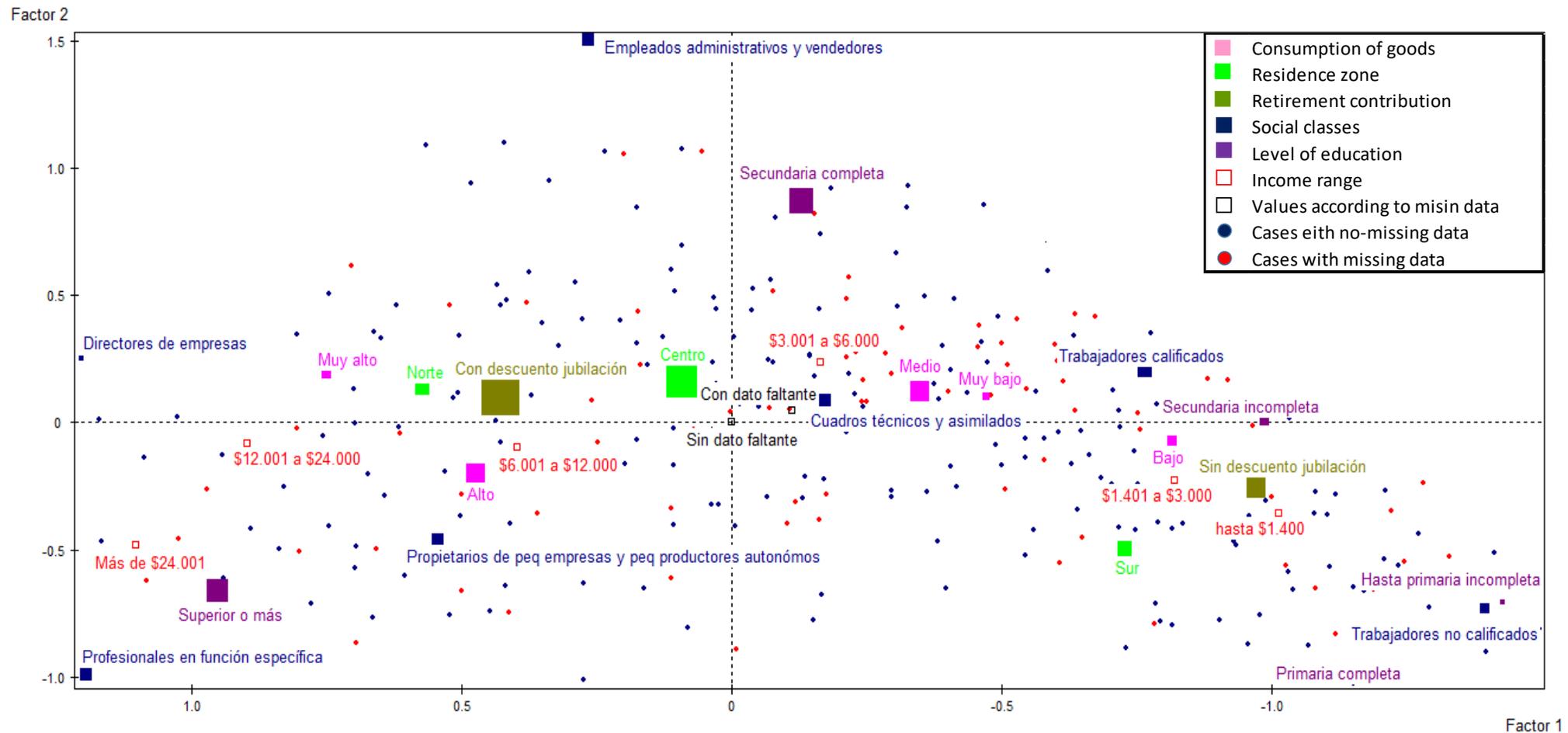
**Table N° 7. Selected variables: own corrected values and percentage of explained inertia.\***

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

Note: \* This percentage was calculated using Greenacre corection (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

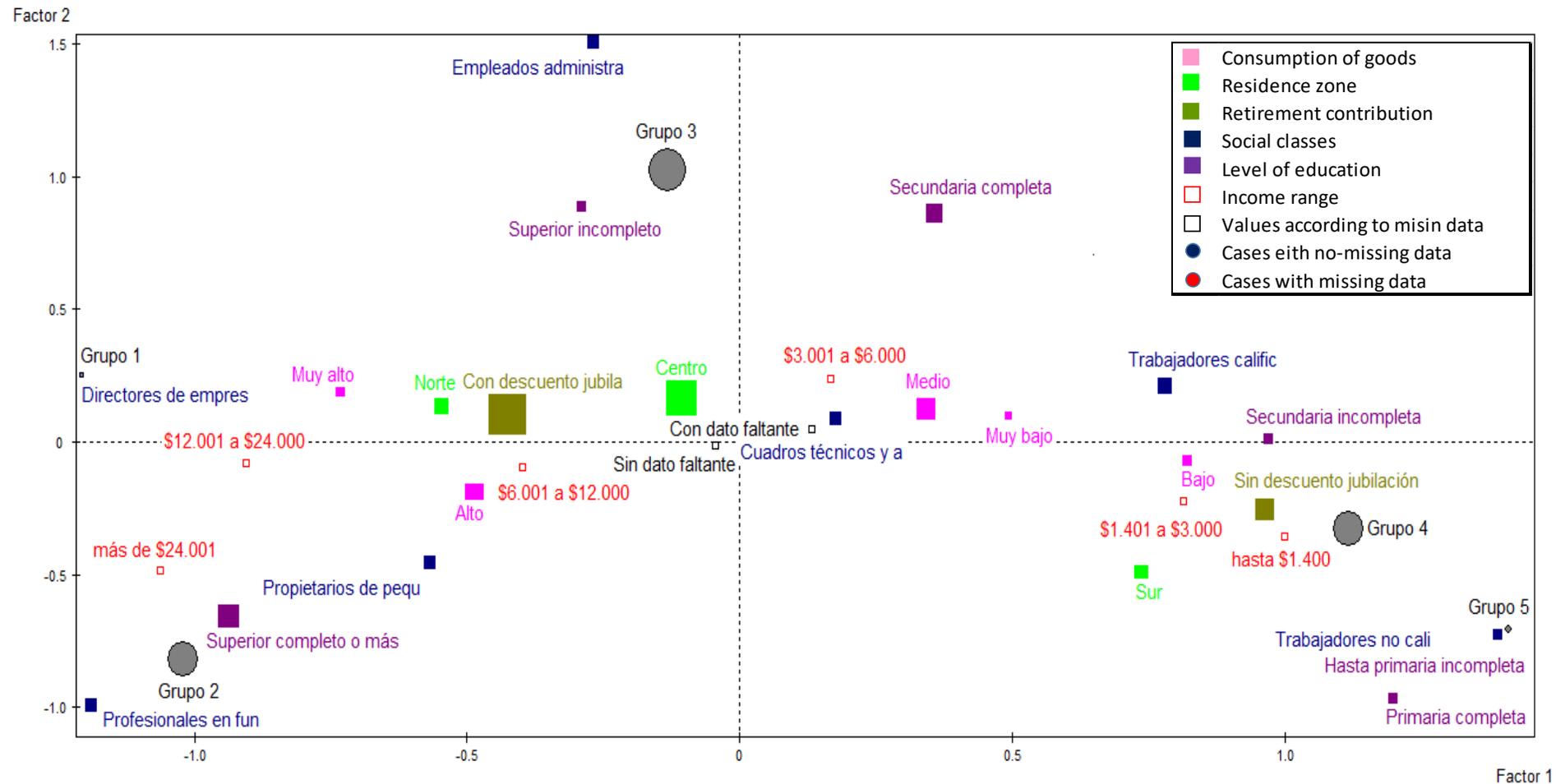
The **first factorial axis** explains 73.6 % of inertia and distributed to the population based on differences in **social class, area of residence , education level and perception of retirement contributions** (which can be thought of as an indicator labor formality).

**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



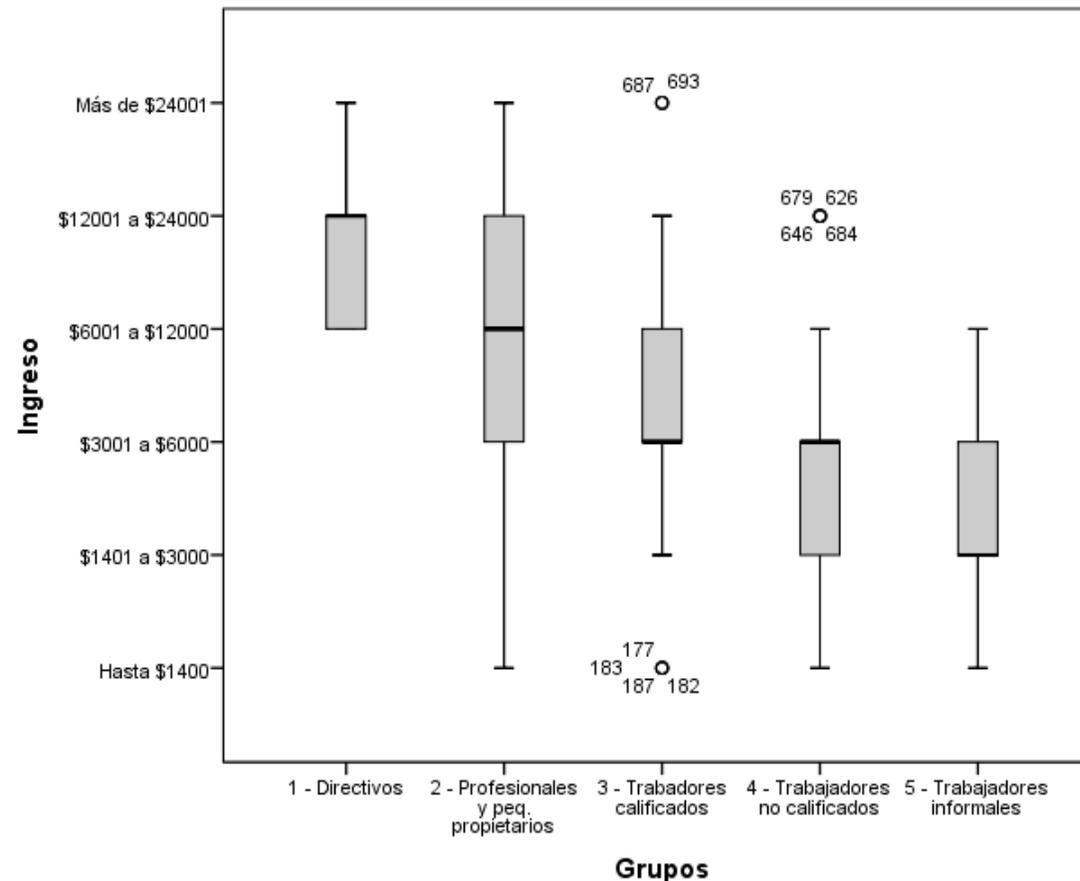
# MULTIPLE CORRESPONDENCE ANALYSIS

## 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).

Source: own elaboration based on Survey FONCYT 2012-2013

# CLUSTERS AND INCOME

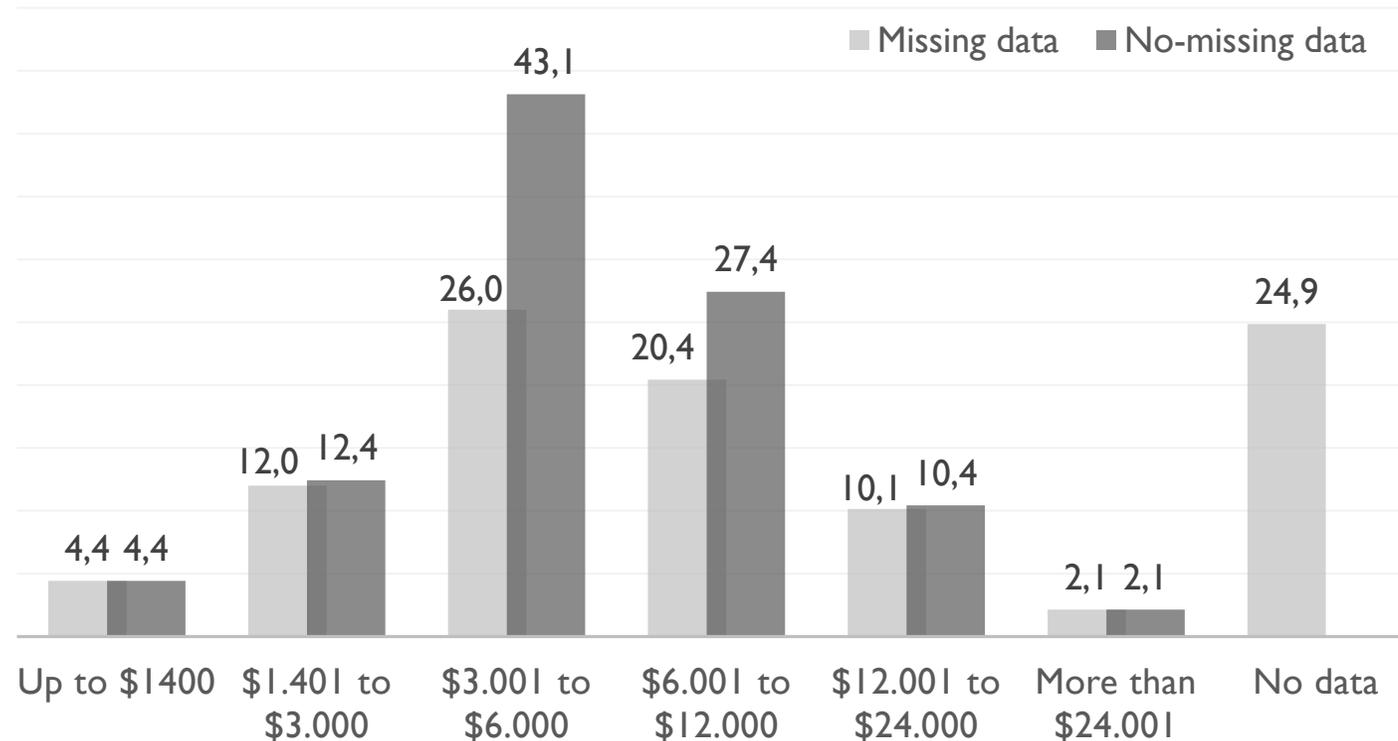
**Table N° 7. Income medians according to clusters**

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

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