



Universitat Autònoma de Barcelona  
Departament de Sociologia

## **ANEXO METODOLÓGICO**

**Nuevo modelo de estratificación social y nuevo  
instrumento para su medición.  
El caso argentino.  
Anexo Metodológico.**

**Tesis Doctoral: Sandra Fachelli Oliva**

**Directores: Jordi Planas i Coll y Pedro López Roldán**

**Barcelona, Octubre 2009**

**Esta tesis se ha realizado con el apoyo del Comissionat per a Universitats i  
Recerca del Departament d'Innovació, Universitats i Empresa de la Generalitat  
de Catalunya y del Fondo Social Europeo.**

## E. Recálculo de Varianza Explicada

Año 1997

Tableau des valeurs propres

Trace de la matrice: 5.33333

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1,000	0,399	7,476	7,476
2,000	0,256	4,803	12,279
3,000	0,231	4,323	16,602
4,000	0,206	3,872	20,474
5,000	0,197	3,699	24,173
6,000	0,193	3,614	27,787
7,000	0,182	3,405	31,192
8,000	0,174	3,265	34,457
9,000	0,171	3,200	37,658
10,000	0,170	3,194	40,852
11,000	0,170	3,181	44,033
12,000	0,169	3,163	47,195
13,000	0,168	3,150	50,345
14,000	0,167	3,135	53,480
15,000	0,166	3,116	56,597
16,000	0,165	3,101	59,697
17,000	0,164	3,080	62,777
18,000	0,164	3,074	65,851
19,000	0,163	3,062	68,913
20,000	0,161	3,025	71,938
21,000	0,156	2,925	74,863
22,000	0,154	2,883	77,747
23,000	0,153	2,874	80,621
24,000	0,149	2,786	83,406
25,000	0,140	2,630	86,036
26,000	0,133	2,501	88,538
27,000	0,127	2,373	90,911
28,000	0,122	2,286	93,196
29,000	0,109	2,049	95,245
30,000	0,104	1,950	97,195
31,000	0,078	1,471	98,666
32,000	0,071	1,334	100,000

Recálculo de Varianza explicada				
Factor	VPT	PVE	PAVE	
1,000	0,232	44,761	44,761	44,761
2,000	0,090	17,266	62,026	62,026
3,000	0,064	12,328	74,354	74,354
4,000	0,040	7,681	82,035	82,035
5,000	0,031	5,909	87,944	87,944
6,000	0,026	5,030	92,973	92,973
7,000	0,015	2,877	95,851	95,851
8,000	0,007	1,444	97,295	97,295
9,000	0,004	0,775	98,070	98,070
10,000	0,004	0,709	98,779	98,779
11,000	0,003	0,576	99,355	99,355
12,000	0,002	0,387	99,742	99,742
13,000	0,001	0,258	100,000	100,000

Año 2002

Tableau des valeurs propres

Trace de la matrice: 4.83333

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1,000	0,384	7,943	7,943
2,000	0,255	5,268	13,211
3,000	0,230	4,750	17,961
4,000	0,203	4,210	22,171
5,000	0,196	4,050	26,221
6,000	0,187	3,871	30,092
7,000	0,176	3,647	33,739
8,000	0,173	3,576	37,315
9,000	0,171	3,532	40,847
10,000	0,169	3,504	44,351
11,000	0,168	3,471	47,822
12,000	0,167	3,455	51,277
13,000	0,167	3,447	54,724
14,000	0,166	3,438	58,162
15,000	0,165	3,412	61,574
16,000	0,163	3,379	64,953
17,000	0,162	3,343	68,296
18,000	0,161	3,325	71,620
19,000	0,157	3,251	74,871
20,000	0,154	3,183	78,055
21,000	0,150	3,102	81,156
22,000	0,147	3,038	84,194
23,000	0,137	2,832	87,026
24,000	0,129	2,678	89,704
25,000	0,126	2,608	92,311
26,000	0,121	2,509	94,820
27,000	0,108	2,241	97,061
28,000	0,074	1,534	98,595
29,000	0,068	1,405	100,000

Recálculo de Varianza explicada				
Factor	VPT	PVE	PAVE	
1,000	0,217	45,412	45,412	45,412
2,000	0,088	18,386	63,798	63,798
3,000	0,063	13,153	76,951	76,951
4,000	0,037	7,700	84,651	84,651
5,000	0,029	6,085	90,736	90,736
6,000	0,020	4,269	95,004	95,004
7,000	0,010	2,011	97,015	97,015
8,000	0,006	1,292	98,307	98,307
9,000	0,004	0,842	99,149	99,149
10,000	0,003	0,563	99,712	99,712
11,000	0,001	0,226	99,937	99,937
12,000	0,000	0,072	100,009	100,009
13,000	0,000	-0,009	100,000	100,000

Año 2003

## Tableau des valeurs propres

Trace de la matrice: 5.16667

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1,000	0,374	7,241	7,241
2,000	0,247	4,772	12,013
3,000	0,235	4,554	16,567
4,000	0,209	4,051	20,618
5,000	0,201	3,886	24,504
6,000	0,192	3,720	28,224
7,000	0,176	3,410	31,634
8,000	0,171	3,313	34,947
9,000	0,171	3,301	38,248
10,000	0,170	3,292	41,540
11,000	0,169	3,267	44,808
12,000	0,168	3,253	48,060
13,000	0,168	3,246	51,307
14,000	0,167	3,239	54,546
15,000	0,167	3,226	57,771
16,000	0,164	3,182	60,953
17,000	0,164	3,167	64,120
18,000	0,163	3,154	67,275
19,000	0,162	3,133	70,408
20,000	0,161	3,110	73,518
21,000	0,158	3,066	76,584
22,000	0,155	2,990	79,574
23,000	0,145	2,811	82,386
24,000	0,140	2,719	85,104
25,000	0,136	2,626	87,731
26,000	0,130	2,525	90,255
27,000	0,123	2,381	92,637
28,000	0,121	2,350	94,987
29,000	0,105	2,029	97,016
30,000	0,085	1,653	98,669
31,000	0,069	1,331	100,000

## Recálculo de Varianza explicada

Factor	VPT	PVE	PAVE
1,000	0,207	42,851	42,851
2,000	0,080	16,495	59,346
3,000	0,069	14,175	73,521
4,000	0,043	8,802	82,324
5,000	0,034	7,050	89,374
6,000	0,026	5,274	94,647
7,000	0,010	1,964	96,612
8,000	0,004	0,929	97,540
9,000	0,004	0,801	98,341
10,000	0,003	0,709	99,050
11,000	0,002	0,445	99,494
12,000	0,001	0,287	99,782
13,000	0,001	0,218	100,000

Año 2006

## Tableau des valeurs propres

Trace de la matrice: 5.00000

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1,000	0,375	7,491	7,491
2,000	0,246	4,925	12,416
3,000	0,224	4,485	16,901
4,000	0,205	4,095	20,995
5,000	0,187	3,731	24,726
6,000	0,173	3,467	28,193
7,000	0,172	3,436	31,629
8,000	0,171	3,425	35,054
9,000	0,170	3,408	38,462
10,000	0,169	3,388	41,850
11,000	0,169	3,379	45,229
12,000	0,168	3,358	48,587
13,000	0,167	3,345	51,932
14,000	0,166	3,328	55,260
15,000	0,166	3,313	58,573
16,000	0,165	3,291	61,863
17,000	0,164	3,287	65,151
18,000	0,163	3,261	68,412
19,000	0,162	3,240	71,651
20,000	0,159	3,171	74,822
21,000	0,158	3,155	77,977
22,000	0,157	3,131	81,108
23,000	0,153	3,067	84,176
24,000	0,139	2,771	86,947
25,000	0,137	2,745	89,691
26,000	0,129	2,577	92,268
27,000	0,124	2,489	94,757
28,000	0,111	2,227	96,984
29,000	0,081	1,620	98,604
30,000	0,070	1,396	100,000

## Recálculo de Varianza explicada

Factor	VPT	PVE	PAVE
1,000	0,208	48,345	48,345
2,000	0,080	18,512	66,856
3,000	0,058	13,395	80,251
4,000	0,038	8,852	89,104
5,000	0,020	4,619	93,723
6,000	0,007	1,556	95,279
7,000	0,005	1,196	96,475
8,000	0,005	1,062	97,537
9,000	0,004	0,872	98,409
10,000	0,003	0,634	99,042
11,000	0,002	0,530	99,572
12,000	0,001	0,288	99,860
13,000	0,001	0,140	100,000

## Anexo N° 5: Diseño de análisis

### A. Indicador ocupación: Variables ilustrativas

Tabla N°1: Distribución de los hogares ocupados según rama de Actividad

Período	Estabilidad	Post Crisis	Recuperación	
			Incipiente	Consolidada
Rama de Actividad	1997	2002	2003	2006
Sector primario	1,2%	1,5%	1,7%	1,2%
Industria electricidad, gas y agua	17,8%	14,0%	15,3%	15,3%
Construcción	9,5%	8,6%	8,5%	10,0%
Comercio	14,7%	14,0%	17,9%	18,0%
Hotelería y transporte	10,5%	10,3%	10,0%	10,4%
Finanzas y activ. Inmobiliarias	9,7%	9,2%	9,9%	10,4%
Administ. pública	8,2%	10,3%	10,1%	8,6%
Enseñanza	6,9%	9,3%	8,8%	8,9%
Salud y servicios sociales	6,6%	6,5%	7,9%	6,2%
Otros Servicios	8,5%	10,5%	6,1%	5,6%
Empleo doméstico	5,5%	5,1%	3,1%	5,1%
Sin especificar	1,0%	0,6%	0,7%	0,2%
Total	100,0%	100,0%	100,0%	100,0%

Tabla N°2: Distribución de los hogares ocupados según sector

Período	Estabilidad	Post Crisis	Recuperación	
			Incipiente	Consolidada
Establecimiento	1997	2002	2003	2006
Público	25,7%	32,8%	29,3%	24,4%
Privado	74,1%	66,6%	69,4%	74,6%
Otro	0,2%	0,6%	1,3%	0,9%
Total	100,0%	100,0%	100,0%	100,0%

### B. Otras pruebas realizadas para analizar vivienda

Para realizar un acercamiento metodológico en los años que abarca nuestra investigación, hemos realizado varias pruebas, analizando la asociación entre diferentes características habitacionales de manera tal de poder identificar diferentes tipos de vivienda. Se presenta a continuación un ejercicio que luego se desestimó debido a que arrojaba resultados similares a los obtenidos de forma más parsimoniosa, es decir, utilizando las variables libremente en el ACM (sin estructurarlas previamente).

Un primer elemento a considerar ha sido el tipo de relación de propiedad de los hogares con respecto a la vivienda, es decir diferenciar aquellos hogares que son propietarios de la vivienda (cerca de un 70% en la década analizada) y aquellos que no lo son. Esta medida, si bien es buena como una primera aproximación, es insuficiente y es necesario mejorarla con algún criterio adicional, pues resulta evidente que la calidad de vivienda, así como la cantidad de miembros que la habitan son elementos imprescindibles para establecer algún tipo de clasificación.

A partir de varios análisis realizados, presentamos a continuación una clasificación, que ha sido el producto de ensayar distintas mediciones por lo que consideramos que ilustra el problema y permite establecer una relación entre lo que se observa en el sector vivienda y su acercamiento empírico. En este indicador intervienen varios elementos, a saber: a) régimen de tenencia de la vivienda, b) tipo de vivienda, c) tenencia y uso de baño y c) hacinamiento.

<b>Tipo de vivienda A</b>	
<b>Regimen de tenencia</b>	Propietario vivienda y terreno o Inquilino
<b>Tipo de vivienda</b>	Casa/departamento
<b>Baño (tenencia y uso)</b>	Baño de uso exclusivo
<b>Hacinamiento</b>	Sin hacinamiento
<b>Tipo de vivienda B</b>	
<b>Hacinamiento</b>	Resto sin hacinamiento
<b>Tipo de vivienda C</b>	
<b>Hacinamiento</b>	Con hacinamiento
<b>Tenencia de baño</b>	Sin baño

*Fuente: elaboración propia*

Así estos tres tipos de vivienda tienen un ordenamiento lexicográfico, según las características de habitabilidad seleccionadas: la vivienda de tipo A es la que posee las mejores condiciones habitacionales (según las variables que hemos seleccionado) e incluye a los propietarios o inquilinos que viven en casa o departamento, que tienen baño de uso exclusivo y que no están hacinados; la vivienda de tipo B se define por exclusión del tipo A, pero por la inclusión de la condición de que no haya hacinamiento. Por último, la vivienda tipo C es la que reúne las condiciones más desfavorables de habitabilidad pues se encuentran todas aquellas viviendas que tienen hacinamiento y/o aquellas que no poseen baño.

Cabe destacar que el concepto “hacinamiento” se refiere a la cantidad de personas por cuarto. Se trata de una medida ampliamente utilizada en nuestro país y hemos considerado, como usualmente se hace, el cociente entre el número de personas en el hogar y la cantidad de habitaciones de uso exclusivo del hogar (distinta de cocina o baño). Se considera que un hogar tiene hacinamiento si dicho valor es superior a dos personas por cuarto.

El cuadro siguiente presenta los resultados obtenidos para los años bajo análisis.

### **Tipo de Vivienda**

*Argentina, 1997-2002-2006*

<b>Vivienda</b>	<b>1997</b>	<b>2002</b>	<b>2006</b>
Vivienda A	60,3	62,4	59,8
Vivienda B	21,5	22,6	27,8
Vivienda C	18,2	14,9	12,4

*Fuente: elaboración propia sobre la base de EPH*

Observamos que el tipo de vivienda que cumple los estándares más altos se mantiene en torno al 60% y aumenta levemente en 2002. El tipo de vivienda intermedia aumenta a través del tiempo, por un efecto de la disminución del conjunto de hogares que se encuentra en peor situación. Estos pasan de representar el 18,2% en 1997 al 12,4% en 2006.

## Anexo Nº 6: Análisis previos a la obtención de los estratos sociales

### A. Síntesis de los Informes de Avance de Investigación Metodológicos Nº3 a Nº6

Comparación entre bases de datos del año 2006

Factor	4to. Trimestre			2do. Semestre		
	VPT	PVE	PAVE	VPT	PVE	PAVE
1	0,290	45,4	45,4	0,280	48,8	48,8
2	0,128	20,0	65,3	0,110	19,3	68,1
3	0,080	12,6	77,9	0,083	14,5	82,6
4	0,051	7,9	85,8	0,043	7,6	90,2
5	0,035	5,5	91,3	0,026	4,6	94,7
6	0,018	2,8	94,2	0,011	2,0	96,7
7	0,010	1,5	95,7	0,007	1,2	97,9
8	0,009	1,4	97,1	0,005	0,8	98,7
9	0,007	1,0	98,1	0,004	0,7	99,3
10	0,006	0,9	99,0	0,003	0,5	99,8
11	0,004	0,6	99,7	0,001	0,1	100,0
12	0,002	0,3	100,0	0,000	0,0	100,0
Total	0,639			0,573		

### Resultados iniciales del ACM

#### Salidas sin ponderar

Particiones			
Seuil 0,5%			
1997	2002	2003	2006
3 factores			
4 y 7	4, 6 y 9	4, 5 y 10	4,5 y 9
4 factores			
3 y 6	5, 8 y 9	5 y 8	5 y 10
5 factores			
6 y 7	3 y 7	6, 9 y 12	5 y 7
6 factores			
5 y 9	3, 6 y 7	3 y 7	6 y 7
7 factores			
4, 8 y 10	8	8 y 9	4 y 7
Seuil 1%			
3 factores			
4, 8 y 9	4, 5 y 8	4, 5 y 7	4, 7 y 10
4 factores			
6 y 9	5, 8 y 9	4, 8 y 9	3, 5 y 6
5 factores			
6 y 8	4, 7 y 10	5 y 7	5 y 8
6 factores			
3 y 6	3, 6 y 7	6 y 7	6
7 factores			
3 y 8	3, 7 y 8	7 y 9	3, 6 y 9

Spad versión 5.0

Varianza Explicada				
Seuil 0,5%				
Factor	1997	2002	2003	2006
1	42,0	42,0	34,2	44,4
2	17,6	17,5	23,2	17,2
3	12,6	12,3	11,9	13,4
4	8,2	7,4	10,4	8,3
Total	80,4	79,2	79,7	83,3
5	6,5	6,8	5,9	5,5
6	5,1	4,9	5,2	3,9
7	3,0	3,1	3,3	2,2
Total 7f	95,0	94,0	94,1	94,9
Seuil 1%				
Factor	1997	2002	2003	2006
1	43,7	43,4	42,6	48,1
2	18,2	18,1	16,3	18,6
3	11,7	11,9	13,7	13,1
4	8,2	7,7	8,6	8,6
Total	81,8	81,1	81,3	88,4
5	6,4	7,0	7,3	4,4
6	5,3	4,9	4,6	2,2
Total	88,1	88,2	88,7	92,8
7	2,7	2,8	2,0	1,4
Total 7f	96,1	95,9	95,3	96,5

Spad versión 5.0

## B. Síntesis de los Informes de Avance de Investigación Metodológicos N°7 y N°8

### 1. Revisión

Se hizo una revisión general, comencé desde cero. Realicé correcciones (excepto educación) pues la variable es distinta en las EPH continuas semestrales.<sup>9</sup>

A continuación se presentan los resultados.

**Salidas sin ponderar: Particiones**

Seuil 1%			
1997	2002	2003	2006
3 factores			
4, 7 y 8	4, 8 y 9	4 y 7	4, 5 y 10
4 factores			
3, 5 y 7	5 y 7	5, 7 y 8	3, 5 y 9
5 factores			
6	6 y 9	6 y 7	3, 6 y 9
6 factores			
6 y 7	3 y 7	6 y 8	4 y 7
7 factores			
7, 8 y 9	7 y 8	4, 6 y 8	4, 8 y 10

*Spad versión 5.0*

Se observa un nivel de homogeneidad muy importante, prácticamente cada factor tiene una clase que coincide.

Seuil 1%				
Factor	1997	2002	2003	2006
1	44,8	45,5	42,8	48,3
2	17,3	18,4	16,5	18,6
3	12,3	13,2	14,2	13,4
<b>Subtot</b>	<b>74,4</b>	<b>77,2</b>	<b>73,4</b>	<b>80,3</b>
4	7,7	7,7	8,8	8,4
<b>Subtot</b>	<b>82,0</b>	<b>84,9</b>	<b>82,2</b>	<b>88,7</b>
5	5,9	6,1	7,0	4,8
<b>Subtot</b>	<b>87,9</b>	<b>91,0</b>	<b>89,2</b>	<b>93,5</b>
6	5,0	4,3	5,3	2,0
7	2,9	2,0	2,0	1,2
<b>Total 7f</b>	<b>95,9</b>	<b>97,3</b>	<b>96,5</b>	<b>96,6</b>

*Spad versión 5.0*

El porcentaje de varianza explicada es relativamente similar según cada factor, excepto el primer factor del año 2006 donde –una vez más– muestra un valor más elevado que el resto.

### 2) ACM con 3 Factores y 4 particiones

El resultado que se presentará a continuación puede considerarse el más homogéneo de todos los análisis realizados hasta el momento aunque sea muy agregado. De alguna manera, este cuadro resume todo lo que se presentará en el resto de este informe.

<sup>9</sup> En una etapa posterior se homogeneizan las variables de educación y se detecta y corrige un error en la no identificación del decil 0 en el año 2006. Se realiza nuevamente el ACM arrojando modificaciones en las particiones y se modifican levemente los porcentajes de los estratos.

Estratos	1997	2002	2003	2006
Estrato alto	15,3	14,0	14,5	15,9
Estrato medio	46,5	43,4	42,5	45,5
Estrato bajo	17,0	20,2	21,7	20,3
Estrato Inactivos	21,2	22,3	21,3	18,3
Total	100,0	100,0	100,0	100,0

Los estratos muestran una relativa homogeneidad en su formación. No obstante, es importante señalar que el “Estrato inactivos” en los años 1997, 2002 y 2003 posee la mayor parte de los hogares desocupados, a diferencia de lo que sucede en 2006 donde los desocupados son menos y se distribuyen entre todas las categorías (aunque el 43% de estos hogares queda comprendido en el “Estrato bajo”).

Esta situación es el principal elemento que estructura la realidad en forma diferente en 2006 con respecto al resto de los años.

Estrato inactivos	1997	2002	2003	2006
Hogares inactivos	71,5	70,2	71,6	73,8
Hogares desocupados	17,1	19,8	14,5	1,7

El “Estrato inactivos” está conformado en más del 70% de hogares inactivos, pero también posee otro tipo de hogares, en su mayor parte desocupados entre 1997 y 2003.

La recuperación económica que comienza lentamente en 2003 se consolida hacia 2006 y este hecho introduce elementos que contribuyen a conformar la realidad de otra manera y en consecuencia los datos reflejan estas modificaciones.

### 3) ACM con 4 Factores y 5 particiones

Hay dos estructuras bien definidas y diferentes entre los años analizados:

3.1) Los años 1997 y 2002 arrojan una estructura social determinada que se presenta a continuación:

Estratos	1997	2002
Estrato alto	14,6	13,9
Estrato medio	43,5	40,2
Estrato bajo	17,1	20,3
Estrato inactivos	19,7	19,1
Estrato desocupados	5,2	6,5
Total	100,0	100,0

Es coherente con el proceso vivido durante esos años, especialmente se refleja el impacto de la crisis entre diciembre de 2001 y febrero de 2002, donde Argentina entró en default, su economía cayó 11% en términos reales e interanuales y la desocupación y la pobreza se dispararon.

3.2) Los años 2003 y 2006 revelan otro tipo de estructuración de la realidad:



Estratos	2003	2006
Estrato alto	12,5	14,0
Estrato medio-alto	22,3	21,3
Estrato medio	31,8	34,2
Estrato bajo	15,3	14,8
Estrato Inactivos*	18,0	15,7
Total	100,0	100,0

*\* En 2003 incluye el remanente de desocupados*

Aquí desaparece el Estrato de desocupados, pues comienza a ser un elemento menos importante. De tal manera que el remanente de desocupados en 2003 se concentra en mayor medida en la categoría que predominan los hogares inactivos (Estrato inactivos).

Ya en 2006 los pocos hogares desocupados que quedan (2,6%) se equidistribuyen entre el estrato medio (37%) y el estrato bajo (37%) y el resto entre las otras categorías.

### C. Síntesis de los Informes de Avance de Investigación Metodológicos N°9 y N°9Bis

Método Hiérarchique Recip. Apliqué esta técnica a todas las bases con 3 y 4 factores y obtuve los siguientes resultados:

#### 1. Particiones y varianza explicada

Método: Hierarchique RECIP					Varianza explicada				
Salidas sin ponderar: Particiones					Factor	1997	2002	2003	2006
<b>Seuil 1%</b>					1	45,3	45,5	43,3	48,6
3 factores					2	17,5	18,4	16,7	18,7
1997	2002	2003	2006		3	12,5	13,2	14,3	13,5
4, 5 y 9					4	7,8	7,7	8,9	8,5
4 y 8					Subtot	75,3	77,2	74,2	80,9
4, 5 y 7					5	6,0	6,1	7,1	4,8
5 y 7					6	5,1	4,3	5,3	2,0
4 factores					7	2,9	2,0	2,0	1,2
5 y 8	5 y 6	4 y 5	3, 5 y 7		Subtot	83,0	84,9	83,1	89,4
Spad versión 5.0					Total 7f	97,0	97,3	97,5	97,4

#### 2. Análisis de las 5 particiones del ACM 4 factores

Del análisis de los 5 grupos, vuelve a surgir los resultados obtenidos en los informes anteriores, es decir que 2006 cambia de estructura: desaparecen los desocupados como grupo en si mismo y se forman dos grupos de trabajadores.

Estratos	1997	Estratos	2006
Estrato alto	14,6	Estrato alto	14,6
Trabajadores	43,4	Trabajadores (formales 70%)	21,3
Estrato bajo	17,1	Trabajadores (formales 50%)	34,6
Estrato desocupados	5,2	Estrato bajo	14,3
Estrato Inactivos	19,7	Estrato Inactivos	15,2
Total	100,0	Total	100,0

Con el objetivo de mirar las diferencias acaecidas, esto es interesante. Pero no así para captar la estructura que deberíamos observar inter-temporalmente. Considero que estas mismas diferencias se pueden captar entrando a analizar el resto de las particiones obtenidas en el Item A del informe N°8.

#### 3. Organización jerárquica y construcción de los grupos

La tabla muestra en 1997 el cruce entre las 5 y 8 particiones. Como se puede observar las agrupaciones no poseen un comportamiento que esperábamos lograr a partir de aplicar la técnica Recip. Pues la idea era ver si se podían obtener agrupaciones que no quedaran distribuidas en pequeños porcentajes (por ejemplo de 1,3% o 0,3%).

Tabla de contingencia Partition en 8 classes \* Partition en 5 classes

Partición en 8 classes		Partition en 5 classes					Total
		Classe 1/5	Classe 2/5	Classe 3/5	Classe 4/5	Classe 5/5	
Classe 1/8	Recuento	1.366.407	60	0	0	0	1.366.467
	% de Partition en 8 classes	100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
	% de Partition en 5 classes	49,5%	0,0%	0,0%	0,0%	0,0%	21,5%
Classe 2/8	Recuento	1.309.829	158.191	0	252.213	0	1.720.233
	% de Partition en 8 classes	76,1%	9,2%	0,0%	14,7%	0,0%	100,0%
	% de Partition en 5 classes	47,4%	12,6%	0,0%	23,2%	0,0%	27,1%
Classe 3/8	Recuento	618	1.075.016	0	0	0	1.075.634
	% de Partition en 8 classes	0,1%	99,9%	0,0%	0,0%	0,0%	100,0%
	% de Partition en 5 classes	0,0%	85,9%	0,0%	0,0%	0,0%	16,9%
Classe 4/8	Recuento	7.294	1.007	243.974	14.308	1.885	268.468
	% de Partition en 8 classes	2,7%	0,4%	90,9%	5,3%	0,7%	100,0%
	% de Partition en 5 classes	0,3%	0,1%	74,0%	1,3%	0,2%	4,2%
Classe 5/8	Recuento	0	0	85.649	0	0	85.649
	% de Partition en 8 classes	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%
	% de Partition en 5 classes	0,0%	0,0%	26,0%	0,0%	0,0%	1,3%
Classe 6/8	Recuento	0	1.020	0	820.065	0	821.085
	% de Partition en 8 classes	0,0%	0,1%	0,0%	99,9%	0,0%	100,0%
	% de Partition en 5 classes	0,0%	0,1%	0,0%	75,5%	0,0%	12,9%
Classe 7/8	Recuento	76.583	15.647	0	101	587.535	679.866
	% de Partition en 8 classes	11,3%	2,3%	0,0%	0,0%	86,4%	100,0%
	% de Partition en 5 classes	2,8%	1,3%	0,0%	0,0%	63,3%	10,7%
Classe 8/8	Recuento	0	0	0	0	338.308	338.308
	% de Partition en 8 classes	0,0%	0,0%	0,0%	0,0%	100,0%	100,0%
	% de Partition en 5 classes	0,0%	0,0%	0,0%	0,0%	36,5%	5,3%
Total	Recuento	2.760.731	1.250.941	329.623	1.086.687	927.728	6.355.710
	% de Partition en 8 classes	43,4%	19,7%	5,2%	17,1%	14,6%	100,0%
	% de Partition en 5 classes	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Se revisa el Informe N°9 que fue realizado el 26-11 por primera vez, a partir de dos hallazgos importantes:

- 1) Las reflexiones volcadas en el Informe N°13
- 2) La detección de un error en la construcción del Decil en el año 2006 (El 0,9% de los ingresos cero quedaron incluidos en la categoría no sabe/no contesta).

La reproducción del año 2006 con el Método Hiérarchique Recip, muestra que también obtengo 4 clases cuando uso tres factores. Esto es muy relevante, pues es posible utilizar esta metodología para llevar a cabo el análisis general.

Particiones y varianza explicada: corrección año 2006

Método: Hierarchique RECIP  
Salidas sin ponderar: Particiones

Seuil 1%			
1997	2002	2003	2006
3 factores			
4, 5 y 9	4 y 8	4, 5 y 7	4, 5 y 8
4 factores			
5 y 8	5 y 6	4 y 5	3, 5 y 7

Spad versión 5.0

Varianza explicada				
Factor	1997	2002	2003	2006
1	45,3	45,5	43,3	48,8
2	17,5	18,4	16,7	18,7
3	12,5	13,2	14,3	13,5
<b>Subtot</b>	<b>75,3</b>	<b>77,2</b>	<b>74,2</b>	<b>81,0</b>
4	7,8	7,7	8,9	8,9
<b>Subtot</b>	<b>83,0</b>	<b>84,9</b>	<b>83,1</b>	<b>90,0</b>
5	6,0	6,1	7,1	4,7
6	5,1	4,3	5,3	1,6
7	2,9	2,0	2,0	1,2
<b>Total 7f</b>	<b>97,0</b>	<b>97,3</b>	<b>97,5</b>	<b>97,4</b>

Las diferencias observadas son:

- a) En el informe anterior las particiones con 3 factores era 5 y 7.
- b) En el informe anterior el porcentaje de varianza explicada del primer factor era 48,6%.

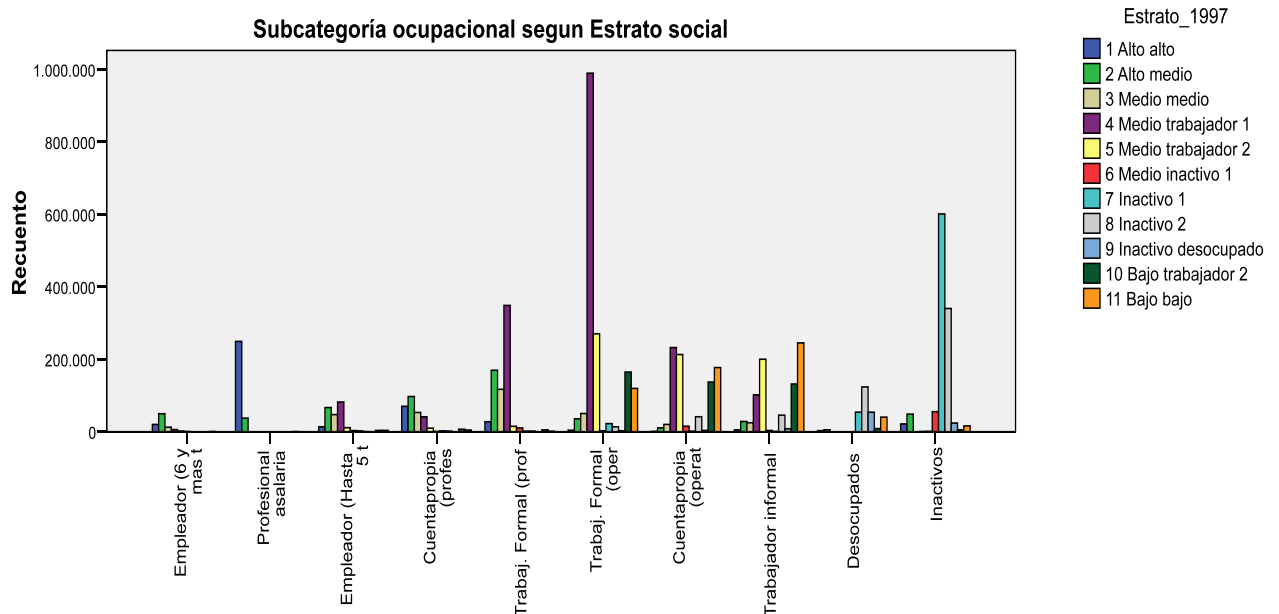
Resultados del ACM de los 4 años utilizando los 3 factores Método RECIP.

Estrato	1997	2002	2003	2006
Alto	14,8	14,0	14,6	16,2
Medio	45,3	43,6	42,5	45,6
Inactivo	21,7	22,4	21,3	17,9
Bajo	18,2	20,0	21,6	20,3
Total	100,0	100,0	100,0	100,0



Estrato_1997	Frecuencia	Porcentaje
1 Alto alto	419.548	6,6
2 Alto medio	551.224	8,7
3 Medio medio	327.709	5,2
4 Medio trabajador 1	1.811.563	28,5
5 Medio trabajador 2	725.433	11,4
6 Medio inactivo 1	91.424	1,4
7 Inactivo 1	686.961	10,8
8 Inactivo 2	570.639	9,0
9 Inactivo desocupado	92.492	1,5
10 Bajo trabajador 2	465.761	7,3
11 Bajo bajo	612.956	9,6
Total	6.355.710	100,0

		Partition en 4 classes				
		Estrato medio	Estrato inactivo	Estrato bajo	Estrato alto	Total
		% del N de la columna	% del N de la columna	% del N de la columna	% del N de la columna	% del N de la columna
Estrato_1997	1 Alto alto	,0%	,0%	,0%	43,2%	6,6%
	2 Alto medio	,0%	,0%	,0%	56,8%	8,7%
	3 Medio medio	11,1%	,0%	,0%	,0%	5,2%
	4 Medio trabajador 1	61,3%	,0%	,0%	,0%	28,5%
	5 Medio trabajador 2	24,5%	,0%	,0%	,0%	11,4%
	6 Medio inactivo 1	3,1%	,0%	,0%	,0%	1,4%
	7 Inactivo 1	,0%	50,9%	,0%	,0%	10,8%
	8 Inactivo 2	,0%	42,3%	,0%	,0%	9,0%
	9 Inactivo desocupado	,0%	6,9%	,0%	,0%	1,5%
	10 Bajo trabajador 2	,0%	,0%	43,2%	,0%	7,3%
	11 Bajo bajo	,0%	,0%	56,8%	,0%	9,6%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%



## E. Síntesis de los Informes de Avance de Investigación Metodológicos N°17 a N°19

### Comparación entre métodos diferentes

Con el fin de analizar el procesamiento realizado por el método RECIP en SPAD y en SPSS, se procede a realizar un ejercicio de comparación y para ello seleccionamos la base de 2002 con menor cantidad de casos.

Resultados comparados (Método RECIP es sin iteración en todos los casos)

SPAD RECIP Estrato Social	Todos los casos 2002	BASE REDUCIDA AÑO 2002			
		Partición	SPAD Inicial*	SPAD RECIP	Quick Cluster
Alto	16,0	Alto	15,8	16,7	37,3
Medio Laboral Activo	44,6	Medio Laboral Activo	45,6	45,3	17,7
Medio Laboral Inactivo	17,6	Medio Laboral Inactivo	17,2	20,2	11,3
Bajo	21,8	Bajo	21,5	17,8	33,7
Total	100	Total	100	100	100

\* Estratos luego de la selección al azar de los 10.000 casos

Fuente: elaboración propia sobre la base de EPH

\* Estratos existentes en la base grande (22.832 casos), pero observados en la base de 10.000 casos

En una primera observación parece que el análisis RECIP se mantiene relativamente estable. No obstante si realizamos un cruce entre la antigua clasificación (en la base de 10.000 casos, que refleja la muestra aleatoria de la base completa) y la nueva clasificación a partir del ACM Recip también sin iterar con los 10000 casos, surgen las siguientes diferencias:

SPAD Inicial 10.000	SPAD RECIP Muestra Reducida				Total
	2 Medio	3 Inactivo	1 Alto	4 Bajo	
1 Alto	0,1	1,7	14,0	0,0	15,8
2 Medio	39,7	1,2	2,7	2,0	45,6
3 Inactivo	0,4	16,7	0,0	0,0	17,2
4 Bajo	5,1	0,6	0,0	15,8	21,5
Total	45,3	20,2	16,7	17,8	100,0

Se presentan los resultados del ACM entre la base completa y la base reducida:

Comparación SPAD		
Factor	Todos los casos	Base Reducida
1	45,5	46,0
2	18,4	18,5
3	13,2	12,4
Total	77,2	76,9
Particiones	4 y 8	4 y 7

A continuación se presentan dos ejercicios con el fin de analizar la sensibilidad de la agrupación de las clases según el método utilizado.

1) **Análisis Quick Cluster** en la base de 10.000 casos con las coordenadas de la base entera del 2002 (22.832 casos).

### Coordenadas de la base grande 2002

Estrato	Eje 1	Eje 2	Eje3
1	0,1271565115	0,2613367110	0,2917917629
2	0,0875984806	0,3845640279	-0,7332978960
3	0,8804440451	-0,7941843420	-0,0485038772
4	-0,7752312187	-0,2862605250	-0,0120760417

**QUICK CLUSTER**

```
Eje1 TO Eje3
/MISSING=LISTWISE
/CRITERIA= CLUSTER(4) MXITER(50) CONVERGE(0)
/METHOD=KMEANS(NOUPDATE)
/PRINT INITIAL ANOVA
/INITIAL= (0.1271565115 0.2613367110 0.2917917629
0.0875984806 0.3845640279 -0.7332978960
0.8804440451 -0.7941843420 -0.0485038772
-0.7752312187 -0.2862605250 -0.0120760417)
/OUTFILE='C:\Documents and Settings\Administrador\Escritorio\8va etapa\2002\3 factores\Muestra 10000
casos\Centres QC-Noupdate.sav'
/SAVE=CLUSTER (QCLW4) DISTANCE (DQCLW4).
```

1.A) Quick cluster y SPAD Inicial: se observa que existe bastante coincidencia entre ambos métodos.

**Tabla de contingencia Estrato: Alto medio inactivo y bajo \* Número inicial de casos**

			Número inicial de casos				Total
			1	2	3	4	
Estrato: Alto medio inactivo y bajo	1 Alto	% de Estrato: Alto medio	88,6%	6,5%	4,7%	0,2%	100,0%
		% de Número inicial de	97,8%	5,0%	1,7%	0,1%	15,8%
	2 Medio	% de Estrato: Alto medio	0,6%	0,7%	93,0%	5,7%	100,0%
		% de Número inicial de	2,0%	1,6%	95,6%	12,3%	45,6%
	3 Inactivo	% de Estrato: Alto medio	0,1%	97,3%	2,5%		100,0%
		% de Número inicial de	0,1%	81,6%	1,0%		17,2%
	4 Bajo	% de Estrato: Alto medio		11,2%	3,7%	85,1%	100,0%
		% de Número inicial de		11,8%	1,8%	87,5%	21,5%
Total	% de Estrato: Alto medio	14,3%	20,5%	44,4%	20,9%	100,0%	
	% de Número inicial de	100,0%	100,0%	100,0%	100,0%	100,0%	

1.B) Quick cluster y SPAD RECIP (reelaboración de estratos en la base de 10.000): se observan coincidencias, pero algo menores que las observadas en 1.A.

**Tabla de contingencia Partition en 4 classes Spad RECIP \* Número inicial de casos**

			Número inicial de casos				Total
			1	2	3	4	
Partition en 4 classes Spad RECIP	Classe 1/4	% de Partition en 4 classes		3,4%	87,5%	9,0%	100,0%
		% de Número inicial de		7,6%	89,3%	19,6%	45,3%
	Classe 2/4	% de Partition en 4 classes	3,4%	89,7%	6,8%	0,0%	100,0%
		% de Número inicial de	4,8%	88,7%	3,1%	0,0%	20,2%
	Classe 3/4	% de Partition en 4 classes	81,3%		18,2%	0	100,0%
		% de Número inicial de	95,2%		6,9%	0	16,7%
	Classe 4/4	% de Partition en 4 classes		4,2%	1,7%	94,1%	100,0%
		% de Número inicial de		3,7%	0,7%	80,1%	17,8%
Total	% de Partition en 4 classes	14,3%	20,5%	44,4%	20,9%	100,0%	
	% de Número inicial de	100,0%	100,0%	100,0%	100,0%	100,0%	

**2) Two Step Cluster análisis** en la base de 10.000 casos con los 3 ejes nuevos (correspondientes al ACM realizado sobre los 10.000 casos)

```
TWOSTEP CLUSTER
/CONTINUOUS VARIABLES = Eje1 Eje2 Eje3
/DISTANCE EUCLIDEAN
/NUMCLUSTERS FIXED = 4
/HANDLENOISE 0
/MEMALLOCATE 100
/CRITERIA INITHRESHOLD (0) MXBRANCH (8) MXLEVEL (3)
/PLOT PIEFREQ
/PRINT COUNT SUMMARY
```



/SAVE VARIABLE=TSC\_1739 .  
 AIM TSC\_1739  
 /CONTINUOUS Eje1 Eje2 Eje3  
 /PLOT ERRORBAR CLUSTER (TYPE=PIE) .

La observación general es que no es un buen método para diferenciar las clases pues agrupa casi el 80% de los casos en una sola categoría.

### 2.A) Two Step Cluster y SPAD Inicial

Tabla de contingencia Estrato: Alto medio inactivo y bajo \* Número de conglomerados en dos fases

			Número de conglomerados en dos fases				Total
			1	2	3	4	
Estrato: Alto medio inactivo y bajo	1 Alto	% de Estrato: Alto	1,3%		36,2%	62,5%	100,0%
		% de Número de	2,3%		7,2%	99,6%	15,8%
	2 Medio	% de Estrato: Alto	0,0%		99,9%	0,1%	100,0%
		% de Número de	0,1%		57,1%	0,3%	45,6%
	3 Inactivo	% de Estrato: Alto	46,9%	1,0%	52,0%	0	100,0%
		% de Número de	89,4%	13,2%	11,2%	0	17,2%
	4 Bajo	% de Estrato: Alto	0	5,4%	91,2%		100,0%
		% de Número de	0	86,8%	24,6%		21,5%
	Total	% de Estrato: Alto	9,0%	1,3%	79,8%	9,9%	100,0%
		% de Número de	100,0%	100,0%	100,0%	100,0%	100,0%

### 2.B) Two Step Cluster y SPAD RECIP

Tabla de contingencia Partition en 4 classes Spad RECIP \* Número de conglomerados en dos fases

			Número de conglomerados en dos fases				Total
			1	2	3	4	
Partition en 4 classes Spad RECIP	Clase 1/4	% de Partition en 4			100,0%		100,0%
		% de Número de			56,7%		45,3%
	Clase 2/4	% de Partition en 4	43,5%	0,9%	55,1%	0,5%	100,0%
		% de Número de	97,8%	13,1%	14,0%	1,1%	20,2%
	Clase 3/4	% de Partition en 4			41,4%	58,6%	100,0%
		% de Número de			8,7%	98,9%	16,7%
	Clase 4/4	% de Partition en 4	0	6,6%	92,3%		100,0%
		% de Número de	0	86,9%	20,6%		17,8%
	Total	% de Partition en 4	9,0%	1,3%	79,8%	9,9%	100,0%
		% de Número de	100,0%	100,0%	100,0%	100,0%	100,0%

### 2.C) Two Step Cluster y Quick Cluster

Tabla de contingencia Número inicial de casos \* Número de conglomerados en dos fases

			Número de conglomerados en dos fases				Total
			1	2	3	4	
Número inicial de casos	1	% de Número inicial de			30,7%	69,3%	100,0%
		% de Número de			5,5%	100,0%	14,3%
	2	% de Número inicial de	43,4%	0,9%	55,7%		100,0%
		% de Número de	98,8%	13,2%	14,3%		20,5%
	3	% de Número inicial de			100,0%		100,0%
		% de Número de			55,6%		44,4%
	4	% de Número inicial de	0,5%	5,6%	93,9%		100,0%
		% de Número de	1,2%	86,8%	24,6%		20,9%
	Total	% de Número inicial de	9,0%	1,3%	79,8%	9,9%	100,0%
		% de Número de	100,0%	100,0%	100,0%	100,0%	100,0%

**3) Two Step Cluster análisis** en la base de 10.000 casos con los 3 ejes viejos (correspondientes al ACM realizado sobre los 22.832 casos.)

TWOSTEP CLUSTER

```

/CONTINUOUS VARIABLES = v052 v053 v054
/DISTANCE EUCLIDEAN
/NUMCLUSTERS FIXED = 4
/HANDLENOISE 0
/MEMALLOCATE 100
/CRITERIA INITHRESHOLD (0) MXBRANCH (8) MXLEVEL (3)
/PLOT BARFREQ PIEFREQ
/PRINT COUNT SUMMARY
/SAVE VARIABLE=TSC_2129 .
AIM TSC_2129
/CONTINUOUS v052 v053 v054
/PLOT ERRORBAR CLUSTER (TYPE=PIE) .

```

Aquí sucede lo mismo que en el caso anterior, por ello no se realizan los cruces, pues el 85% de los casos aparecen agrupados en el conglomerado 3.

**Número de conglomerados en dos fases**

		Frecuencia	Porcentaje	válido	acumulado
Válidos	1	125.083	4,0	4,0	4,0
	2	232.945	7,5	7,5	11,6
	3	2.622.968	84,9	84,9	96,5
	4	108.870	3,5	3,5	100,0
	Total	3.089.866	100,0	100,0	