

Análisis del nivel de error en el Censo Nacional de Población, Hogares y Viviendas 2001. Descripción e interpretación de resultados.

Alvarez, Gustavo Oscar.

Cita:

Alvarez, Gustavo Oscar (2003). *Análisis del nivel de error en el Censo Nacional de Población, Hogares y Viviendas 2001. Descripción e interpretación de resultados. VII Jornadas Argentinas de Estudios de Población. Asociación de Estudios de Población de la Argentina, Tafí del Valle, Tucumán.*

Dirección estable: <https://www.aacademica.org/gustavo.oscar.alvarez/37>

ARK: <https://n2t.net/ark:/13683/pq6q/YeW>



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>.

ANÁLISIS DEL NIVEL DE ERROR EN EL CENSO NACIONAL DE POBLACIÓN, HOGARES Y VIVIENDAS 2001. DESCRIPCIÓN E INTERPRETACIÓN DE RESULTADOS *

*Gustavo O. Álvarez ***

RESUMEN

Este trabajo tiene como primer objetivo describir la calidad de los datos del Censo Nacional de Población, Hogares y Vivienda 2001, en cuanto al “nivel de error” correspondiente a las distintas variables. Al respecto, se considerarán dos expresiones del mismo: la “no respuesta” y la “multimarca”.

El antecedente principal de este análisis es la revisión sistemática del nivel de error que se efectuó con los datos del Censo Nacional de Población de Hogares y Vivienda 1991.

En tal sentido, se reeditará una evaluación de las variaciones en el “nivel de error” detectado en las distintas jurisdicciones en que se llevó a cabo el relevamiento censal. Este examen se orientará en general a identificar el cumplimiento de los umbrales de tolerancia definidos a priori, y en particular a diferenciar la concentración geográfica de los errores de relevamiento.

Por otra parte, este trabajo apunta a interpretar los factores asociados al “nivel de error”. Al respecto, resulta de particular interés examinar la relación entre las variables afectadas por los mayores niveles de error. Complementariamente, se establecerán comparaciones entre el último censo y el anterior, a fin de identificar tendencias en la calidad de los datos. Este ejercicio permitirá establecer la eficacia de las acciones intercensales.

A fin de cumplir con los propósitos de este trabajo, se utilizará una combinación de técnicas de análisis multivariado. En principio, se aplicará el análisis factorial de componentes principales para describir las múltiples correlaciones entre los niveles de error atribuibles a distintas variables. Posteriormente se procederá a estratificar las jurisdicciones mediante el análisis de clusters provisto por la técnica de k-medias.

* Este trabajo se realizó en el marco de las acciones del Programa de Evaluación de Calidad del Censo Nacional de Población, Hogares y Viviendas 2001, dirigidas por Gladys Massé, Directora de Estadísticas Poblacionales, bajo la coordinación de María Cecilia Rodríguez Gauna y Mariela Goldberg.

** Dirección de Estadísticas Poblacionales, INDEC , galva@indec.mecon.gov.ar

El autor desea agradecer los comentarios de Alejandro Giusti, Gladys Massé, Mariela Goldberg, Alicia Gómez, Eduardo Movia y Diana Munilla, así como las tareas de asistencia llevadas a cabo por Ariel Matto y Diego Ventrici.

1. Introducción

La utilidad de la información derivada de las fuentes de datos sociodemográficos depende en gran medida de la calidad con que han sido elaborados los datos. Esta consideración general es aplicable al último Censo Nacional de Población, Hogares y Viviendas que se relevó en la Argentina a mediados de noviembre del 2001. Al respecto, resulta significativo evaluar la calidad de aquel operativo reconociendo sus hechos más salientes, las líneas generales y aspectos particulares. Asimismo la comparación con el relevamiento censal precedente, efectuado en 1991, puede dar cuenta de los logros y las dificultades que se dieron en su diseño conceptual.

Entre las múltiples enfoques que pueden hacerse de los errores¹, este documento se concentra en los errores manifiestos en los datos obtenidos a nivel de cada registro. En tal sentido, se examinarán las falencias que se expresan como 'blancos' y 'multimarcas' en la información que se produce durante el relevamiento a través de un examen indirecto sin apelar a una vuelta al terreno². Por los resultados del control de calidad durante el proceso de lectura de las cédulas censales, es posible descartar el aporte de errores atribuibles a imperfecciones del procesamiento de los datos durante su captura y asumir que las fallas se dieron en el trabajo de campo.

En particular, esta presentación se refiere al funcionamiento de las variables relevadas a nivel de los hogares, esto es aquellos contenidos que fueron comunes al grupo de personas que vivían bajo un mismo techo y compartían los gastos de alimentación. En otros instrumentos de relevamiento, el Censo 2001 relevó características globales de los segmentos y atributos individuales que no serán objeto de este estudio.

La calidad de los datos será examinada en referencia al 'nivel del error' (en adelante ER) por respuestas ausentes o inespecíficas. Entre las primeras se identifican las fallas por 'No Respuesta' (NR). El porcentaje de NR se obtiene por la relación numérica entre los casos 'sin marca' y el total de marcas esperadas de acuerdo al flujo del cuestionario. En tanto que se reconocen dos tipos de respuestas inespecíficas: los errores por 'Marca Múltiple' (MM) y las respuestas correspondientes a la categoría 'Ignorado'. El porcentaje de MM se calcula como el cociente entre casos con 'más de una marca'³ sobre el total de marcas esperadas. En definitiva, el error total de cada variable -ER- es la suma de NR más MM e 'Ignorado'.

A efectos de hacer una presentación más ágil de los resultados las variables serán referidas mediante siglas (que se explicitan en el Cuadro 1) al tiempo que se representará a las áreas de estudio a través de abreviaturas que se detallan a continuación:

¹ Un examen exhaustivo de criterios de clasificación de errores en los operativos censales se encuentra en Massé, 1997.

² Ver la distinción entre procedimientos de evaluación directa e indirecta en Chackiel, J y Macció, G., 1978.

³ Ninguna pregunta del Censo 2001 admitía más que una marca válida.

Tabla 1. Referencia de Áreas Analizadas

Jurisdicciones y áreas de estudio	Referencia	Código
Total del País	TOT	01
Ciudad de Buenos Aires	CIU	02
Provincia de Buenos Aires	BAI	07
Partidos del Gran Buenos Aires	GBA	04
Resto de la Provincia de Buenos Aires	RBA	06
Catamarca	CAT	10
Córdoba	CBA	14
Corrientes	CRN	18
Chaco	CHA	22
Chubut	CHU	26
Entre Ríos	ERI	30
Formosa	FOR	34
Jujuy	JUJ	38
La Pampa	PMP	42
La Rioja	RJA	46
Mendoza	MZA	50
Misiones	MIS	54
Neuquén	NEU	58
Río Negro	RNG	62
Salta	SAL	66
San Juan	SJU	70
San Luis	SLU	74
Santa Cruz	SCR	78
Santa Fe	SFE	82
Santiago del Estero	SGO	86
Tucumán	TUC	90
Tierra del Fuego, Antártida e Islas del Atlántico Sur	TDF	94

2. Error de respuesta en el Censo 2001

2.1 Papel de los defectos por respuesta ausente o múltiple

El Error Total (ER) reúne distintas expresiones defectuosas de los datos obtenidos. En algunas preguntas cierta parte de esa imprecisión se originó en casos donde se tomó la opción 'ignorado', sin embargo este tema sólo afectó a las pocas preguntas de la cédula de hogares que incluyeron la misma⁴. Por tal motivo, la composición del ER se analizará fundamentalmente a través de la participación relativa de los defectos por NR y MM.

En el cuadro 1, se aprecia que en todas las variables ha sido mayor la incidencia de la NR frente a la MM. Al respecto, se encuentra que solamente la variable 'Material de los techos' tiene más de uno por ciento de MM cuando el error de NR supera ese nivel en la mayoría de los rubros.

Por otra parte, cabe consignar que la relación entre ambos factores es heterogénea. En efecto, un grupo de variables presenta una participación bastante pareja de ambas modalidades donde la razón es de menos de tres casos de NR por cada MM. Son las variables de 'Material de los techos', 'Material de las paredes' y 'Provisión de agua'. Otras tienen una participación moderadamente minoritaria ya que cuentan con hasta quince casos de NR por cada MM, son otras variables como 'Condición de Ocupación', 'Combustible para Cocinar', 'Cantidad de Habitaciones para Dormir', 'Cantidad de Habitaciones en Total', 'Material de los pisos', 'Tipo de Vivienda', 'Procedencia del Agua', 'Desagüe del Inodoro', 'Presencia de personas con Discapacidad' y 'Pueblo indígena de pertenencia'. Pero en todas las demás variables, esto es la mayoría, la incidencia de la MM es irrelevante puesto que aparece menos de una cada treinta NR y en el caso más extremo 'Presencia de Descendientes de un Pueblo Indígena' sólo aparecen cada trescientas respuestas ausentes.

En líneas generales, se aprecia que la MM tiene una incidencia mínima y que su ocurrencia es aún menos frecuente en preguntas donde sólo hay dos opciones de respuesta. En efecto, la MM casi no se advierte en las preguntas de saneamiento de la vivienda y posesión de bienes durables que ofrecían respuestas por 'sí' o por 'no'. Contrariamente, eleva su participación –aunque conservando niveles mínimos– en las preguntas por materiales de la vivienda (techo, paredes, pisos) donde el criterio de predominancia no habría funcionado adecuadamente.

⁴ Las únicas preguntas de la cédula de hogar que ofrecían la opción de respuesta 'Ignorado' fueron la referida a 'Presencia de cielorraso/revestimiento interior en el techo' y la de 'Propiedad del Terreno'.

Cuadro 1. Nivel de errores por 'NR' y 'MM' por variables de hogar del Censo 2001, Total del país (en porcentajes)

Variable	Pregunta	Referencia	NR	MM
Condición de Ocupación	S/n	NRE	1,79	0,43
Hogar con descendiente o perteneciente a un pueblo indígena	2	IND	6,92	0,02
A qué pueblo indígena pertenece	2b	PIN	8,14	0,56
En este hogar hay alguna persona con discapacidad	3	DIS	5,01	0,69
Tipo de Vivienda	4	TVI	0,96	0,07
Material de los pisos	5	PIS	1,04	0,19
Material de las paredes	6	PAR	1,02	0,34
Paredes tienen revestimiento externo	7	REV	0,60	0,02
Material de los techos	8	TCH	1,17	1,09
Presencia de cielorraso/revestimiento interior en el techo	9	CIE	1,13	0,04
Provisión de agua	10	TAG	1,65	0,85
Procedencia de agua	11	DAG	1,14	0,17
Tenencia de baño	12	BAN	1,27	0,03
Tenencia de inodoro	13	INO	0,80	0,01
Tenencia de botón/cadena/mochila en el inodoro	14	BOT	2,31	0,01
Desagüe del inodoro	15	DES	0,53	0,06
Exclusividad en el uso del baño	16	EXC	2,30	0,01
Tenencia de lugar para cocinar	17	COC	2,38	0,01
Tenencia de instalación de agua en la cocina	18	ACO	1,48	0,01
Combustible para cocinar	19	CMB	1,87	0,44
Cantidad de habitaciones para dormir	20	HAD	0,78	0,15*
Cantidad de habitaciones en total	21	HAT	0,84	0,16*
Régimen de tenencia	22	REG	1,17	0,03
Propiedad del terreno	23	PRO	1,57	0,01
Tenencia de heladera con freezer/freezer solo	24a	HFR	3,02	0,10
Tenencia de heladera sin freezer	24b	HEL	4,07	0,12
Tenencia de lavarropas automático	24c	LAU	3,48	0,08
Tenencia de lavarropas común	24d	LAC	4,09	0,09
Tenencia de videocasetera/reproductor	24e	VID	2,61	0,06
Tenencia de teléfono celular	24f	TLC	3,15	0,08
Tenencia de teléfono fijo	24g	TLF	2,42	0,06
Tenencia de televisión por cable	24h	TVC	2,27	0,11
Tenencia de horno microondas	24i	HOR	3,12	0,03
Tenencia de computadora con conexión a internet	24j	INT	3,93	0,04
Tenencia de computadora sola	24k	CPU	3,96	0,04

* En estas variables, el valor de MM se refiere a respuestas inespecíficas por defectos en la lectura de los números manuscritos por el censista

Fuente: Elaboración propia sobre evaluación de calidad del Censo 2001

2.2 Nivel de error total: diferencias por rubro y por provincia

El nivel de Error Total (ER tot) expresa en términos globales la calidad de los datos obtenidos a escala nacional, sin embargo este valor no indica las diferencias entre las áreas de estudio. Por ello, esa lectura debe ser complementada por la consideración de la heterogeneidad entre áreas de relevamiento, al tiempo que puede compararse con un nivel de error tolerable definido *a priori*. En el cuadro 2, el 'ER tot' se presenta junto al mínimo (ER min) y máximo (ER max) valor hallados a nivel de las áreas de relevamiento. Complementariamente se indica el área donde se encontró cada uno de esos valores, y el coeficiente de variabilidad (CV). La última columna remite al umbral de error fijado *a priori*. En tal sentido, la referencia utilizada fue el umbral aplicado durante el Censo 1991.

En líneas generales, se advierte que el nivel de ER ha sido bajo en la mayoría de las variables. El promedio nacional no ha superado el umbral de error en ningún caso y en la mayoría de los rubros tampoco lo hizo el máximo valor a nivel de las áreas de estudio.

Los mayores niveles de ER se encontraron en un grupo de variables que no fueron diseñadas para su relevamiento universal sino como insumo para aplicar encuestas complementarias al Censo 2001. Así es que los niveles de ER superiores a cinco por ciento se hallaron en 'Pueblo indígena de pertenencia', 'Pertenencia a un pueblo indígena' y 'Presencia de personas con Discapacidad'. Estos resultados ofrecen un indicio de las dificultades de relevar estas temáticas en un operativo universal y politemático como es un censo, y apoyan la decisión de abordarlos mediante estudios *ad hoc* como la Encuesta Nacional de Discapacidad y la Encuesta Complementaria de Pueblos Indígenas.

Entre las restantes, las variables que tuvieron peor funcionamiento han sido la mayor parte de las referidas a bienes durables, donde se apreció un ER tot entre 3 y 5 por ciento. Tal fue el caso de las respuestas referidas a tenencia de heladera (con y sin freezer), lavarropas (común y automático), computadora (con y sin conexión a internet), teléfono celular y horno microondas. Si bien cada una de estas cuestiones se indagó por separado, es posible que haya habido un efecto de irradiación por el cual parte de los errores de NR en 'Tenencia de heladera sin freezer' y 'Tenencia de lavarropas común' se hayan generado como sobreentendidos después de haber respuestas positivas en las anteriores.

Opuestamente, el mejor nivel de calidad se obtuvo en algunas variables de condiciones habitacionales referidas al saneamiento (Tenencia de Inodoro, Desagüe del Inodoro), a los materiales (Presencia de Revestimiento Externo en las Paredes) y al tamaño (Cantidad de Habitaciones en Total y Cantidad de Habitaciones para Dormir).

Cuadro 2. Nivel de Error Total (ER) promedio nacional, mínimo y máximo en variables de hogar del Censo 2001 (en porcentajes)

Variable	ER tot	ER min	Min	ER max	Max	CV	Umbral
Condición de Ocupación	2,23	0,96	CBA	3,57	GBA	28,90	5,0
Hogar con descendiente o perteneciente a un pueblo indígena	6,94	3,67	SCR	10,53	CRN	30,16	
A qué pueblo indígena pertenece	8,70	3,61	NEU	28,53	CRN	60,94	
En este hogar hay alguna persona con discapacidad	5,70	3,16	SCR	7,03	SGO	20,97	
Tipo de Vivienda	1,03	0,45	PMP	1,47	SAL	23,69	3,0
Material de los pisos	1,23	0,60	PMP	1,76	SAL	22,51	3,0
Material de las paredes	1,36	0,61	PMP	2,21	SAL	25,03	3,0
Paredes tienen revestimiento externo	0,62	0,36	PMP	0,96	SGO	21,85	3,0
Material de los techos	2,26	1,15	TDF	3,04	SAL	23,25	3,0
Presencia de cielorraso/revestimiento interior en el techo	1,34	0,59	PMP	2,10	SGO	26,00	3,0
Provisión de agua	2,50	1,54	PMP	6,35	SGO	38,34	3,0
Procedencia de agua	1,31	0,83	SCR	2,48	SGO	30,38	3,0
Tenencia de baño	1,30	0,53	PMP	2,13	SAL	29,90	3,0
Tenencia de inodoro	0,82	0,35	SCR	3,57	MIS	70,02	5,0
Tenencia de botón/cadena/mochila en el inodoro	2,32	1,74	PMP	3,62	SGO	20,76	5,0
Desagüe del inodoro	0,59	0,35	PMP	1,17	SGO	32,16	5,0
Exclusividad en el uso del baño	2,31	1,27	CIU	3,87	SGO	27,86	5,0
Tenencia de lugar para cocinar	2,39	1,30	PMP	3,61	SGO	22,57	5,0
Tenencia de instalación de agua en la cocina	1,50	0,96	PMP	2,44	SGO	26,43	5,0
Combustible para cocinar	2,31	1,18	SCR	4,39	SGO	35,18	5,0
Cantidad de habitaciones para dormir	0,93	0,53	PMP	1,63	SAL	29,29	5,0
Cantidad de habitaciones en total	1,00	0,59	PMP	1,66	SAL	27,10	5,0
Régimen de tenencia	1,19	0,64	PMP	2,19	SGO	29,77	10,0
Propiedad del terreno	1,94	1,21	PMP	2,98	CIU	23,61	5,0
Tenencia de heladera con freezer/freezer solo	3,12	1,54	PMP	4,74	SGO	25,41	5,0
Tenencia de heladera sin freezer	4,20	1,43	PMP	6,79	RJA	31,69	5,0
Tenencia de lavarropas automático	3,56	1,45	PMP	6,46	SGO	37,10	5,0
Tenencia de lavarropas común	4,18	1,81	PMP	6,08	SGO	23,56	5,0
Tenencia de videocasetera/reproductor	2,67	1,22	PMP	5,39	SGO	37,06	5,0
Tenencia de teléfono celular	3,24	1,37	PMP	6,01	SGO	33,50	5,0
Tenencia de teléfono fijo	2,48	1,12	PMP	5,41	SGO	38,92	5,0
Tenencia de televisión por cable	2,38	0,99	PMP	4,65	SGO	35,76	5,0
Tenencia de horno microondas	3,15	1,29	PMP	6,01	SGO	36,32	5,0
Tenencia de computadora con conexión a internet	3,97	1,75	PMP	6,70	SGO	29,99	5,0
Tenencia de computadora sola	4,00	1,62	PMP	6,24	SGO	27,51	5,0

Fuente: Elaboración propia sobre evaluación de calidad del Censo 2001

El valor mínimo de error (ER min) observado entre las áreas en estudio en algunas variables es menos de la quinta parte del umbral y en tres variables está por debajo de la décima parte. En tal sentido, se destacan los ER min de 'Régimen de Tenencia', 'Tenencia de Inodoro' y 'Desagüe del Inodoro'.

En cuanto a las áreas de estudio que tuvieron los rendimientos extremos se ve una marcada regularidad. Al respecto, el nivel de ER en La Pampa fue el mínimo observado en la mayoría de las variables, siendo también frecuente la obtención de errores mínimos en Santa Cruz. En tanto que el máximo nivel de error (ER max) se encontró frecuentemente en dos provincias: Santiago del Estero y Salta. Sistemáticamente el peor nivel de calidad en los datos referidos a tenencia de bienes durables se encontró en Santiago del Estero, así como la peor captación de los materiales predominantes de la vivienda en Salta.

Por último, las distintas variables presentan semejantes grados de heterogeneidad entre las áreas de estudio. El coeficiente de variabilidad (CV) se encuentra entre un veinte y un cuarenta por ciento del valor medio. Sin embargo, dos variables se destacan por su heterogeneidad: la 'Tenencia de Inodoro' y el 'Pueblo indígena de Pertenencia' con valores de CV superiores a sesenta. En el caso de la primer variable, se explica por un nivel de error muy elevado en Misiones; en efecto, es el único caso con ER superior a 3,5 cuando las áreas que le siguen no llegan al 2,5 de ER. La heterogeneidad de los niveles de calidad en 'Pueblo indígena de Pertenencia' son atribuibles a tres áreas donde el ER supera el veinte por ciento (Corrientes, Santiago del Estero y Tierra del Fuego).

3. Evolución del Error entre 1991 y 2001

3.1 Evolución de las respuestas ausentes y las marcas múltiples

Sobre la base de los datos anteriores a la consistencia, en la década del '90 se hizo un análisis del nivel de error del Censo Nacional de Población y Vivienda 1991 (Censo 1991)⁵. Por tal motivo, resulta de gran relevancia compararlo con los resultados actuales a efectos de describir la evolución del nivel de error entre ambos recuentos.

Este ejercicio de comparación entre los dos últimos censos es posible pero deben establecerse algunas condiciones y limitaciones de este análisis. En primer lugar, hubo cambios en el diseño de las variables pero también se dieron modificaciones técnicas en la instrucción de los censistas, en la organización del relevamiento y en la lectura de las cédulas censales, al tiempo que no puede desconocerse las diferencias de contexto político y social entre ambos relevamientos. Por tal motivo, el análisis de los avances y retrocesos de la calidad de los registros obtenidos no admite la identificación específica de cada factor sino apreciar el efecto combinado de ellos.

Por otra parte, una importante cantidad de variables indagadas en el Censo 2001 no lo fueron en el precedente. Esta apreciación cobra mayor relevancia si se tiene en cuenta que en aquel conjunto de 'variables nuevas' se encuentran justamente las que peor rendimiento tuvieron en el último censo (hogares de referencia de encuestas complementarias y tenencia de bienes durables). Asimismo hubo contenidos del Censo 1991 que no se volvieron a relevar ('Cantidad de Hogares' y 'Tenencia de Electricidad').

Otra diferencia entre ambos relevamientos radica en que durante el Censo 1991 se aplicaron dos modelos de cédula: la 'A' (Ampliada) y la 'B' (Básica). En localidades más pobladas se combinó la aplicación a una muestra de la cédula 'A' y al resto la cédula 'B', mientras que en las áreas rurales y localidades menos pobladas se utilizó exclusivamente la 'A'. Esto dio lugar a dos series de NR y MM que en general presentan escaso margen de diferencia. La comparación que sigue fue realizada con los resultados obtenidos por la cédula 'A' ya que tuvo aplicación en todas las provincias y localidades del país⁶.

⁵ Giusti, 1997.

⁶ La información publicada permite reconstruir el ER obtenido conjuntamente con ambas cédulas en lo que se denomina variables básicas, pero no hay publicado un tratamiento semejante de las NR y MM. Ver Giusti, 1997.

Como ya se señaló, el nivel de error total (ER) suma las fallas debidas a ausencia de respuesta (NR) y a las marcas múltiples (MM). En tal sentido, se planteó la comparación entre ambos censos en cuanto a la participación relativa de los defectos por NR y MM.

En ambos censos hubo un predominio de los errores por NR frente a la MM según lo indica el cuadro 3. Sin embargo, en el Censo 1991 hubo una variable en la que se había dado mayor frecuencia de MM que de NR y fue en la ‘Condición de Ocupación’. El diseño del censo anterior incluía en la pregunta 1 un ítem común para reconocer las viviendas ocupadas y los motivos de no realización de entrevista. A propósito de esta variable, en el análisis de calidad del Censo 1991 se planteaban las causas de este error y se concluía sobre la necesidad de:

“realizar una revisión conceptual y operacional en el futuro, no sólo en lo que hace a las diferencias entre vivienda ocupada/desocupada y habitada/deshabitada sino por el tratamiento especial que se dio a los casos de viviendas con moradores temporalmente ausentes” (Giusti, 1997; 26).

Otra variable donde hubo un marcado declive en el nivel de MM entre ambos censos fue en ‘Tenencia de Inodoro’. Este error que ya había sido detectado como un problema de diseño en el Censo 1991 puesto que la pregunta 12⁷ indagaba simultáneamente por la disponibilidad de inodoro con descarga y por la exclusividad en el uso del baño. Al respecto, se juzgó la necesidad de distinguir con mayor precisión los casos de ausencia de baño o de ausencia de inodoro. Consecuente con esta proposición, en el Censo 2001 se desglosó esta indagación en una batería de preguntas por ‘sí’ o ‘no’ con el mencionado descenso en el nivel de error (tanto de MM como de NR).

Cuadro 3. Nivel de errores por ‘NR’ y ‘MM’ por variables de hogar según Censo 2001 y Censo 1991, Total del país (en porcentajes)

Variable	Referencia	NR		MM	
		Censo 2001	Censo 1991	Censo 2001	Censo 1991
Condición de Ocupación	NRE	1,79	1,40	0,43	2,90
Tipo de Vivienda	TVI	0,96	1,20	0,07	0,10
Material de los pisos	PIS	1,04	0,80	0,19	0,10
Material de las paredes	PAR	1,02	1,00	0,34	0,10
Material de los techos	TCH	1,17	1,30	1,09	0,20
Provisión de agua	TAG	1,65	2,20	0,85	0,10
Procedencia de agua	DAG	1,14	1,30	0,17	0,10
Tenencia de inodoro	INO	0,80	2,80	0,01	1,40
Desagüe del inodoro	DES	0,53	2,70	0,06	0,10
Combustible para cocinar	CMB	1,87	3,10	0,44	0,20
Cantidad de habitaciones en total	HAT	0,84	2,00	0,16	0,10
Régimen de tenencia	REG	1,17	6,60	0,03	0,10

Fuente: Giusti, 1997 y elaboración propia sobre evaluación de calidad del Censo 2001

Entre ambos censos, hubo otras variables que tuvieron un incremento de las MM tal que se acercaron a la frecuencia de las respectivas NR. Esta fue la evolución de las variables de ‘Material de los techos’, ‘Material de las paredes’, ‘Material de los Pisos’, ‘Provisión de agua’, ‘Procedencia de agua’ y ‘Combustible para cocinar’.

⁷ En la cédula ‘A’; en la cédula ‘B’ se trataba de la pregunta 8.

Con todo, no se modificó entre ambos censos dos características definitorias de la MM: la incidencia mínima y su relevancia menor frente a la NR.

3.2 Evolución del nivel de error por rubro

El Error Total (ER tot) a nivel nacional también puede compararse en cuanto a su evolución entre los últimos censos. De manera semejante, en el cuadro 4, se presenta la comparación de los indicadores de la heterogeneidad entre áreas de estudio (Er min, Min, ER max, Max, CV) al tiempo que se confronta con el umbral de error.

Cuadro 4. Nivel de Error Total (ER) promedio nacional, mínimo y máximo en variables de hogar según Censo 2001 y Censo 1991 (en porcentajes)

Variable	ER tot		ER min		Min		ER max		Max		CV		Umbral
	2001	1991	2001	1991	2001	1991	2001	1991	2001	1991	2001	1991	
Condición de Ocupación	2,2	4,5	1,0	3,0	CBA	MZA	3,6	6,9	GBA	RBA	29	27	5,0
Tipo de Vivienda	1,0	0,8	0,4	0,4	PMP	CIU	1,5	1,7	SAL	JUJ	24	35	3,0
Material de los pisos	1,2	1,2	0,6	0,4	PMP	CHU	1,8	2,0	SAL	SAL	23	33	3,0
Material de las paredes	1,4	1,2	0,6	0,4	PMP	MZA	2,2	2,1	SAL	RBA	25	26	3,0
Material de los techos	2,3	1,5	1,1	0,6	TDF	MZA	3,0	2,5	SAL	SAL	23	21	3,0
Provisión de agua	2,5	1,7	1,5	0,7	PMP	CIU	6,3	5,2	SGO	FOR SGO	38	90	3,0
Procedencia de agua	1,3	1,2	0,8	0,5	SCR	CHU	2,5	3,4	SGO	JUJ SAL	30	60	3,0
Tenencia de inodoro	0,8	3,4	0,4	1,3	SCR	CIU	3,6	6,8	MIS	SGO	70	45	5,0
Desagüe del inodoro	0,6	2,4	0,3	0,5	PMP	CIU	1,2	6,8	SGO	CBA	32	69	5,0
Combustible para cocinar	2,3	3,3	1,2	2,1	SCR	CBA	4,4	5,3	SGO	SAL	35	16	5,0
Cantidad de habitaciones en total	1,0	1,8	0,6	0,9	PMP	CIU	1,7	3,4	SAL	CAT	27	39	5,0
Régimen de tenencia	1,2	6,1	0,6	3,1	PMP	CIU	2,2	9,8	SGO	CBA	30	25	10,0

Fuente: Giusti, 1997 y elaboración propia sobre evaluación de calidad del Censo 2001

En términos globales, se reconoce un descenso del nivel de ER durante el último censo. Al respecto, se encuentra que esta tendencia se dio en seis variables mientras que sólo en cinco hubo incrementos del ER. Asimismo la magnitud de los cambios ha sido más apreciable entre las primeras que en el segundo grupo de variables.

La evolución más favorable se produjo principalmente en 'Régimen de Tenencia', 'Tenencia de Inodoro' y 'Condición de Ocupación' con más de dos puntos porcentuales de descenso. También hubo mejoras por descenso del nivel de ER en 'Desagüe del Inodoro', 'Combustible para cocinar' y 'Cantidad de habitaciones en total'.

Opuestamente hubo un moderado incremento del nivel de ER en las variables de materiales de la vivienda (material predominante de los techos y de las paredes) y en la 'Provisión del Agua'. Asimismo aumentó el nivel de ER en la variable 'Tipo de Vivienda'.

Los mencionados deterioros de la calidad merecen un examen detenido, aún cuando fueron de nivel moderado. El incremento del ER en 'Material de los techos' puede atribuirse a cambios en la presentación de las opciones de respuesta ya que en el Censo 2001 se excluyó la categoría 'Ignorado'. En cambio, no puede explicarse el mayor nivel de ER de 'Provisión de Agua' desde el diseño, ya que en ambos censos se presentó de manera idéntica la pregunta y las opciones de respuesta.

Las variables que no alteraron el nivel de ER tot han sido 'Material de los Pisos' y 'Procedencia del Agua'. Sin embargo, los niveles nacionales semejantes han sido promedios de situaciones variadas. En tal sentido, hubo mayor homogeneidad entre las provincias en el Censo 2001, el ER min de ambas variables aumentó al tiempo que hubo un descenso del ER max, resultando en importantes declives del CV entre ambas fechas.

En la mayoría de las variables se dio un aumento de la homogeneidad en el nivel de ER entre las áreas de estudio. Esto fue expresión por una parte, de aumento en el nivel de ER min registrado en varios rubros y del descenso del ER max obtenido en casi todos los contenidos. Resultado de ello fue que la comparación entre áreas de estudio en el Censo 2001 registra una menor variabilidad (CV) en la mayoría de las variables y es particularmente marcada para aspectos vinculados con el saneamiento (provisión y procedencia del agua, desagüe del inodoro).

Hubo cambios en el Censo 2001 en cuanto a las áreas donde se verificaron los mayores y menores niveles de ER. Mientras en el Censo 2001 la mayor parte de los ER min provienen de La Pampa y Santa Cruz, en el censo precedente tenían los mínimos Ciudad de Buenos Aires, Chubut y Mendoza. En tanto que la concentración del ER max en dos provincias (Salta y Santiago del Estero) advertida en el Censo 2001 no tiene equivalente en el relevamiento anterior.

El análisis de las áreas en que se produjeron los mayores niveles de error señala que hay dos variables referidas a materiales de las viviendas (de los pisos y de los techos) que tuvieron su rendimiento más débil en la misma área (Salta) y de manera semejante, ocurrió con la 'Provisión del Agua' en Santiago del Estero. Otro caso para profundizar en nuevos estudios, es el error de la 'Condición de Ocupación' que tuvo sus peores niveles en la misma provincia (Buenos Aires) aunque en dos zonas sociodemográficamente diferentes; en el Censo 2001 la peor captación de esta variable se dio en los partidos de Gran Buenos Aires, mientras que previamente se había producido en el Resto de Buenos Aires.

La última consideración acerca de la evolución entre el Censo 1991 y el Censo 2001, se refiere a la relación entre los resultados y los respectivos umbrales. En tal sentido, el dato más saliente es que en el Censo 2001 el umbral es superado –a nivel de las áreas de estudio– sólo en dos variables ('Material de los Techos' y 'Provisión de Agua') mientras que en el censo anterior, había seis variables donde ocurrió lo propio ('Provisión de Agua', 'Procedencia del Agua', 'Tenencia de Inodoro', 'Desagüe del Inodoro', 'Combustible para cocinar' y 'Condición de ocupación de la vivienda'). Por otra parte, el comportamiento de algunas variables en el Censo 2001 ofrece respaldo para ajustar los umbrales a niveles de mayor exigencia.

4. El Nivel de Error y sus componentes principales

El tercer propósito de este trabajo es reconocer patrones de comportamiento en los errores por variable a fin de clasificar a las áreas de estudio. Como se vio en los resultados anteriores, el nivel de error nacional es un promedio de situaciones con un considerable grado de variabilidad.

Dado que el nivel de error cometido en una variable puede estar correlacionado con el ocurrido en las demás, se hizo un análisis exploratorio de las asociaciones entre las frecuencias de errores cometidos. A tal fin, se construyó una matriz de correlaciones entre los ER de las distintas variables hallados a nivel de cada área de estudio⁸.

⁸ La matriz de correlaciones completa entre el nivel de error de las variables relevadas se encuentra en el cuadro I del Anexo Estadístico.

La complejidad de esa matriz multidimensional puede reducirse a un conjunto de factores que reflejan de manera resumida el comportamiento de las variables. En tal sentido, se aplicó el análisis factorial de componentes principales para identificar las dimensiones implícitas más relevantes que contenían a las treinta y cinco variables examinadas. Dado el nivel alto de interrelación entre las variables, se halló que cinco factores resumían en gran medida la variabilidad original de los datos⁹. En el cuadro 5, se representa el grado de correlación entre cada variable y cada uno de los cinco componentes principales¹⁰.

⁹ Entre los cinco componentes principales, explicaban el 91,4% de la varianza total. Una síntesis del análisis factorial se encuentra en el Cuadro II del Anexo Estadístico.

¹⁰ El criterio para identificar factores principales fue que los autovalores (*eigenvalues*) superasen la unidad.

Cuadro 5. Matriz de componentes principales¹¹

	Componente				
	1	2	3	4	5
INO	0,9150	0,1220	0,0016	0,1110	-0,0226
DES	0,8980	0,2360	0,1440	0,1060	0,1170
ACO	0,8640	0,2630	0,3300	0,1740	0,0680
CMB	0,8190	0,2690	0,2550	0,0866	0,0496
BOT	0,8100	0,3580	0,2850	0,2040	0,0950
EXC	0,7930	0,1510	0,4300	0,1870	0,1270
HAD	0,7550	0,4520	0,3860	0,0335	-0,0163
HAT	0,7430	0,4660	0,4180	-0,0305	0,0208
TLF	0,7170	0,3300	0,5600	0,1460	0,1450
TVC	0,7080	0,4030	0,5400	0,1060	0,0732
REG	0,7070	0,4830	0,4320	0,0767	0,1150
COC	0,7010	0,5540	0,3880	0,1260	0,0804
HOR	0,6920	0,3680	0,5770	0,1400	0,0943
TAG	0,6750	0,2430	0,4810	0,0931	0,1610
VID	0,6720	0,3100	0,6140	0,2060	0,1310
DAG	0,6370	0,5610	0,3360	0,0601	0,2210
CIE	0,3970	0,8100	0,3830	0,1350	0,0574
PIS	0,3860	0,7980	0,3980	0,1290	-0,0909
PRO	0,2080	0,7700	0,0061	0,1950	0,2060
TVI	0,2850	0,7540	0,3610	-0,0921	-0,2610
LAC	0,0654	0,7330	0,5610	0,2330	0,0465
PAR	0,5290	0,6850	0,2530	0,0366	-0,1670
REV	0,4970	0,6690	0,1670	0,4100	0,1090
BAN	0,6470	0,6670	0,3020	0,0485	-0,0367
CPU	0,4960	0,5770	0,5730	0,1460	0,0587
HFR	0,4240	0,2760	0,8130	0,1880	0,0510
TCH	0,0782	0,5430	0,7480	-0,1440	-0,0112
LAU	0,5690	0,2770	0,7120	0,2020	0,0897
DIS	0,4250	0,2340	0,6630	0,4230	-0,2000
TLC	0,6140	0,3790	0,6540	0,1730	0,0941
INT	0,5730	0,4270	0,6500	0,1300	0,0918
IND	0,4630	0,1330	0,6360	0,5260	0,0114
HEL	0,2580	0,6090	0,6230	0,1950	0,0552
PIN	0,1430	0,1830	0,1930	0,9050	-0,0058
NRE	-0,2170	-0,0007	-0,0721	0,0148	-0,9160

Fuente: Elaboración propia sobre evaluación de calidad del Censo 2001

¹¹ El cuadro 5 presenta la correlación entre las variables individuales y los componentes principales luego de la rotación efectuada con la técnica *varimax*.

Los valores resaltados en el cuadro 5 indican cuál es valor más importante que cada variable tiene en sus correlaciones con los cinco componentes principales. Entonces se encuentra que las variables referidas a condiciones sanitarias y del uso de la vivienda (tales como tenencia de inodoro, agua en la cocina y cantidad de habitaciones) tienen una relación más marcada con el primer componente. En cambio, el segundo componente es al que más se asocian las características constructivas de las viviendas (materiales predominantes y tipo de vivienda). El tercer componente principal, concita los mayores niveles de correlación con la presencia de ciertos bienes durables (Heladera con o sin freezer, Lavarropas Automático, Teléfono Celular). Por último, los componentes cuarto y quinto se extraen al efecto de dar cuenta de variables con comportamiento singular: 'Pueblo Indígena de Pertenencia' y 'Condición de Ocupación'.

Sobre la base de estos resultados del análisis exploratorio, se construyó un conjunto de índices de nivel de error referidos a las cinco dimensiones implícitas. A tal efecto, se trataron separadamente las variables correspondientes a cada dimensión y se identificó el respectivo componente principal. Por lo tanto, cada índice es una nueva variable generada por la regresión lineal múltiple a partir de todas las variables consideradas, bajo el requisito de que la variación explicada sea la máxima posible¹². En la tabla 2 se presenta un esquema.

Tabla 2. Índices Construidos

Índice	Dimensión	Variabes incluidas	% de Varianza Explicada
Indice_M	Nivel de Error en Materiales y Características Constructivas de la Vivienda	TVI, PIS, PAR, REV, TCH, CIE	81,5
Indice_E	Nivel de Error en Insumos para Encuestas Complementarias	IND, PIN, DIS	81,8
Indice_S	Nivel de Error en Condiciones Sanitarias y del Uso de la Vivienda	TAG, DAG, BAN, INO, BOT, DES, EXC, COC, ACO, CMB, HAD, HAT, REG, PRO	80,9
Indice_B	Nivel de Error en Tenencia de Bienes Durables	HFR, HEL, LAU, LAC, VID, TLC, TLF, TVC, HOR, INT, CPU	90,0
NRE_Z	Nivel de Error Estandarizado en Condición de Ocupación	NRE	No corresponde

En los primeros cuatro casos, el índice ha sido el puntaje factorial del principal componente. La unidimensionalidad de los rubros combinados en cada uno se aprecia por los elevados porcentajes de varianza explicada. Mientras que el ER en 'Condición de Ocupación' se representó a través de la estandarización de esa única variable¹³.

Estas distintas dimensiones, permiten reconocer bloques dentro de la cédula aplicada a hogares que tuvieron comportamiento asociado en cuanto al nivel de error. Asimismo también hay interrelaciones entre estos bloques que se verifican en la matriz de correlaciones de los índices construidos que se expone en el cuadro 6.

¹² Con un valor de r^2 superior a cualquier otro hallable por medio de otras ecuaciones con los datos analizados. Para mayores detalles sobre la técnica de análisis factorial de componentes principales ver Comrey, 1985.

¹³ Con media aritmética igual a cero y desvío típico igual a uno.

Cuadro 6. Matriz de correlaciones entre índices construidos de nivel de error, Total del País, Censo 2001

	Índice_M	Índice_E	Índice_S	Índice_B	NRE_z
Índice_M	1,000	0,629	0,822	0,855	-0,093
Índice_E	0,630	1,000	0,675	0,777	-0,081
Índice_S	0,822	0,675	1,000	0,905	-0,278
Índice_B	0,855	0,777	0,905	1,000	-0,230
NRE_z	-0,093	-0,081	-0,278	-0,230	1,000

Fuente: Elaboración propia sobre evaluación de calidad del Censo 2001

De la correlación entre los índices construidos, se deduce que hay asociación entre el nivel de error de los contenidos descriptivos de las viviendas (características constructivas y condiciones sanitarias) con los rubros que indagan sobre tenencia de bienes durables. En cambio, la calidad de las respuestas en insumos para encuestas complementarias se asocia débilmente con aquellos contenidos y la calidad de la variable 'Condición de Ocupación' tiene un comportamiento muy distinta a las demás.

5. Estratos de áreas según calidad de datos

Los índices de nivel de error construidos a partir del análisis de componentes principales fueron utilizados para clasificar a las áreas de estudio en estratos. El procedimiento utilizado para asignar las unidades de análisis a los estratos fue el método de las *k-medias*.

El método de aglomeración de las *k-medias* aglutina las unidades en un número *k* de grupos que es definido a priori. Para ello se define una cantidad *k* de puntos centrales o centroides; el grupo de pertenencia de cada unidad se define por el centro con el cual se halla la menor distancia. Cuando todos los individuos han sido agrupados en *k* conglomerados, se calculan los centros para cada uno de los grupos. A partir de estos centros de conglomerados, se reagrupan los individuos obteniendo un nuevo agrupamiento¹⁴.

El procedimiento de aglomeración identificó cuatro estratos según comportamiento del nivel de error en las variables de hogar. En el cuadro 7 se sintetiza este resultado presentando las áreas en estudio según el estrato de pertenencia y el puntaje obtenido en cada índice de calidad de los datos.

¹⁴ Ferrán Aranaz, M. (1996) Cap. 17.

Cuadro 7. Áreas de estudio por estrato de pertenencia y puntaje obtenido en índices de nivel de error, Censo 2001

Área	Estrato de Pertenencia	Índice _M	Índice _E	Índice _S	Índice _B	NRE_z
PMP	E1	-2,21	-1,40	-1,62	-1,85	-0,72
SCR	E1	-1,41	-1,57	-1,30	-1,27	-0,31

CBA	E2	-0,75	-0,12	-0,59	-0,55	-1,72
RNG	E2	-1,22	-1,30	-0,93	-0,91	<u>0,79</u>
MZA	E2	-0,44	-0,57	-0,96	-0,84	-0,37
CHU	E2	-0,70	-1,04	-0,87	-0,58	0,41
NEU	E2	-1,00	-1,11	-0,96	-0,95	<u>1,89</u>
CIU	E2	<u>0,57</u>	-0,69	-0,61	-0,84	0,09
TDF	E2	-0,15	-0,24	-0,90	-1,12	<u>0,93</u>
RBA	E2	-0,79	0,39	-0,80	-0,45	<u>0,76</u>
ERI	E2	-0,52	0,11	-0,17	-0,19	-0,05
SLU	E2	-0,23	-0,07	-0,28	-0,05	0,26
SFE	E2	-0,43	<u>1,03</u>	-0,36	-0,03	0,32

SJU	E3	<u>0,85</u>	-0,19	0,03	-0,21	-0,68
CHA	E3	0,02	-0,30	<u>0,82</u>	0,22	-0,78
JUJ	E3	<u>0,58</u>	-1,17	<u>0,81</u>	0,25	-0,05
TUC	E3	0,41	0,27	0,41	<u>1,23</u>	-1,42
MIS	E3	-0,12	<u>0,89</u>	<u>1,00</u>	0,14	0,07
FOR	E3	0,24	0,23	<u>1,32</u>	0,35	-0,02
CAT	E3	<u>0,75</u>	<u>0,50</u>	<u>0,64</u>	<u>0,81</u>	-0,44

RJA	E4	<u>0,75</u>	<u>1,32</u>	0,42	<u>1,09</u>	-0,71
GBA	E4	<u>0,96</u>	<u>0,69</u>	-0,07	<u>0,64</u>	<u>2,78</u>
SAL	E4	<u>2,11</u>	0,24	<u>1,49</u>	<u>1,49</u>	-0,07
CRN	E4	<u>0,96</u>	<u>2,28</u>	<u>1,14</u>	<u>1,20</u>	<u>0,60</u>
SGO	E4	<u>1,78</u>	<u>1,83</u>	<u>2,34</u>	<u>2,44</u>	-1,54

Fuente: Elaboración propia sobre evaluación de calidad del Censo 2001

Cada uno de los estratos constituidos presenta características distintivas:

el estrato E1 incluye a las provincias con los menores puntajes en error. Son provincias donde el nivel de error ha sido bajo en todas las dimensiones temáticas y las que frecuentemente presentaron los mínimos valores de error: La Pampa y Santa Cruz;

el estrato E2 contiene áreas en las que los puntajes de error fueron relativamente bajos en la mayoría de las dimensiones, sin embargo varias de ellas tuvieron niveles de error importantes en la variable 'Condición de Ocupación'. Con variaciones internas, suelen tener niveles de error inferiores al promedio en el índice de calidad de 'Condiciones Sanitarias y Uso de la Vivienda': Córdoba, Río Negro, Mendoza, Chubut, Neuquén, Ciudad de Buenos Aires, Tierra del Fuego, Resto de Buenos Aires, Entre Ríos, San Luis y Santa Fe;

el estrato E3 se compone por provincias con niveles de error relativamente altos en casi todas las dimensiones aunque presentan bajos niveles de error en la variable 'Condición de Ocupación'. Son las provincias en que la mayoría tiene altos niveles de error en las 'Condiciones Sanitarias y Uso de la Vivienda': San Juan, Chaco, Jujuy, Tucumán, Misiones, Formosa y Catamarca;

el estrato E4 tiene las áreas con los mayores niveles de error, particularmente en las características constructivas y la tenencia de bienes durables. Un caso atípico en este grupo es el de Gran Buenos Aires ya que en algunos aspectos se aproxima al estrato anterior aunque se termina diferenciando por el alto nivel de error en 'Condición de Ocupación'. En general, estas son las áreas que presentan los máximos niveles de error: Salta, Santiago del Estero, Corrientes, La Rioja y Gran Buenos Aires.

En definitiva, el comportamiento que primó para clasificar a las áreas fue el nivel de error combinado en los cuatro índices agregados y el resultado obtenido en la 'Condición de Ocupación'.

6. Conclusiones

El balance general de los resultados analizados permite indicar que el nivel de error de las variables de hogar en el Censo 2001 ha sido bajo. En su composición primó la NR, en tanto que la MM tuvo una manifestación marginal. Este último tipo de falla ha sido aún menos frecuente en preguntas donde había opciones dicotómicas (por 'sí' o por 'no') y tiene una participación más apreciable en las que se indagó por el criterio de predominancia que no habría sido interpretado en forma adecuada.

Los ER en el total nacional han sido aceptables ya que en ningún rubro han superado el umbral de tolerancia. Los contenidos con mayores niveles de ER fueron los que se vincularon con temas que no tenían tradición en el relevamiento censal: por una parte, los insumos para el relevamiento de las encuestas complementarias a poblaciones específicas (discapacitados y pertenecientes a pueblos indígenas) y por otra, la indagación por la existencia de bienes durables en el hogar. Con todo, entre estos últimos aspectos el ER no fue tan elevado.

En cuanto a la evolución entre el Censo 1991 y el Censo 2001, se encuentra que el nivel de error por MM se redujo de manera considerable en dos rubros que fueron rediseñados a la luz del análisis de errores posterior al penúltimo censo ('Condición de Ocupación' y 'Tenencia de Inodoro'). Mientras que el nivel de ER descendió en la mayoría de las variables de hogares, al tiempo que se dio una mayor homogeneidad entre los valores provinciales. Las variables que tuvieron la evolución más favorable han sido las relacionadas con las condiciones sanitarias y el uso de las viviendas. Contrariamente, se aprecia un moderado incremento del ER en las variables referidas a características constructivas de la vivienda que ameritan una revisión para el rediseño del futuro censo.

El análisis del nivel de ER de las variables de hogar indica la existencia de un alto grado de interrelación entre estas situaciones. En efecto, la calidad de las treinta y cinco variables examinadas puede ser resumida en cinco componentes principales, según el error en:

- Materiales y Características Constructivas de la Vivienda;
- Condiciones Sanitarias y del Uso de la Vivienda;
- Tenencia de Bienes Durables;
- Insumos para Encuestas Complementarias;
- Condición de Ocupación de la Vivienda.

Asimismo es destacable un importante nivel de asociación entre los tres primeros aspectos. Opuestamente tiene un comportamiento independiente de las demás la 'Condición de Ocupación de la Vivienda'.

En cuanto a la exploración de patrones de ocurrencia de error, se identificó cuatro estratos de jurisdicciones. El grupo de las provincias mejor posicionadas (La Pampa y Santa Cruz) se distingue netamente de las áreas que tuvieron mayores niveles de ER (Gran Buenos Aires, La Rioja, Corrientes, Salta y Santiago del Estero). En tanto que los otros dos estratos presentan situaciones intermedias más cercanas a uno u otro extremo,

pero además diferenciadas por el tipo de errores ya que en uno hubo fallas en 'Condición de Ocupación' mientras que en el restante fueron más frecuente los ER en las 'Condiciones Sanitarias y del Uso de la Vivienda'.

A modo de reflexión final, se puede señalar que el carácter exploratorio de este análisis debería reforzarse por nuevas indagaciones orientadas a comprobar el efecto del diseño conceptual en los errores. No obstante ello, estos resultados marcan un punto de partida para discriminar los factores invalidantes y orientar el diseño del futuro censo.

Referencias bibliográficas

- COMREY, A., (1985) **Manual de análisis factorial**, Madrid, Cátedra.
- CHACKIEL, J. y MACCIO, G. (1978), **Evaluación y corrección de datos demográficos**, Serie B, n° 39, Santiago de Chile, CELADE,.
- FERRAN ARAÑAZ, M. (1996), **SPSS para Windows. Programación y análisis estadístico**. Serie Mac Graw-Hill de Informática, Madrid, Mac Graw-Hill-Interamericana.
- GIUSTI, A. (1997), "Los errores en el Censo de 1991. Primeros avances sobre su magnitud" en **Censo Nacional de Población y Vivienda 1991 Serie J n° 2 Evaluación de la calidad de datos y avances metodológicos 1° parte**, Buenos Aires, INDEC.
- MASSE, G. (1997), "Aspectos conceptuales relativos a la evaluación de la calidad del censo" en **Censo Nacional de Población y Vivienda 1991 Serie J n° 2 Evaluación de la calidad de datos y avances metodológicos 1° parte**, Buenos Aires, INDEC.
- UNITED NATIONS (1999), **Handbook on population and housing census editing**. New York, ST/ESA/STAT/SER.F/82.
- WELTI, C. ed. (1997), **Demografía I**, México, CELADE-PROLAP-UNAM.

ANEXO ESTADISTICO

Cuadro I. Matriz de correlaciones

	NRE	IND	PIN	DIS	TVI	PIS	PAR	REV	TCH	CIE	TAG	DAG	BAN	INO	BOT	DES	EXC
NRE	1,00	-0,18	-0,08	0,04	0,06	-0,06	-0,05	-0,20	-0,08	-0,17	-0,29	-0,33	-0,14	-0,20	-0,26	-0,26	-0,36
IND	-0,18	1,00	0,66	0,94	0,44	0,61	0,50	0,61	0,55	0,61	0,70	0,66	0,65	0,51	0,73	0,59	0,74
PIN	-0,08	0,66	1,00	0,56	0,20	0,43	0,34	0,58	0,14	0,41	0,33	0,28	0,31	0,24	0,42	0,26	0,41
DIS	0,04	0,94	0,56	1,00	0,58	0,67	0,56	0,63	0,64	0,66	0,66	0,66	0,68	0,48	0,67	0,58	0,68
TVI	0,06	0,44	0,20	0,58	1,00	0,89	0,83	0,63	0,75	0,84	0,47	0,63	0,80	0,38	0,53	0,44	0,53
PIS	-0,06	0,61	0,43	0,67	0,89	1,00	0,90	0,82	0,74	0,97	0,66	0,81	0,91	0,48	0,73	0,57	0,63
PAR	-0,05	0,50	0,34	0,56	0,83	0,90	1,00	0,77	0,53	0,86	0,62	0,74	0,88	0,54	0,72	0,60	0,66
REV	-0,20	0,61	0,58	0,63	0,63	0,82	0,77	1,00	0,43	0,87	0,63	0,79	0,82	0,55	0,73	0,69	0,67
TCH	-0,08	0,55	0,14	0,64	0,75	0,74	0,53	0,43	1,00	0,74	0,53	0,59	0,65	0,17	0,46	0,29	0,41
CIE	-0,17	0,61	0,41	0,66	0,84	0,97	0,86	0,87	0,74	1,00	0,69	0,86	0,92	0,47	0,75	0,64	0,64
TAG	-0,29	0,70	0,33	0,66	0,47	0,66	0,62	0,63	0,53	0,69	1,00	0,85	0,71	0,66	0,85	0,76	0,82
DAG	-0,33	0,66	0,28	0,66	0,63	0,81	0,74	0,79	0,59	0,86	0,85	1,00	0,89	0,61	0,88	0,82	0,74
BAN	-0,14	0,65	0,31	0,68	0,80	0,91	0,88	0,82	0,65	0,92	0,71	0,89	1,00	0,67	0,87	0,76	0,72
INO	-0,20	0,51	0,24	0,48	0,38	0,48	0,54	0,55	0,17	0,47	0,66	0,61	0,67	1,00	0,80	0,84	0,79
BOT	-0,26	0,73	0,42	0,67	0,53	0,73	0,72	0,73	0,46	0,75	0,85	0,88	0,87	0,80	1,00	0,88	0,86
DES	-0,26	0,59	0,26	0,58	0,44	0,57	0,60	0,69	0,29	0,64	0,76	0,82	0,76	0,84	0,88	1,00	0,84
EXC	-0,36	0,74	0,41	0,68	0,53	0,63	0,66	0,67	0,41	0,64	0,82	0,74	0,72	0,79	0,86	0,84	1,00
COC	-0,26	0,73	0,39	0,73	0,74	0,87	0,83	0,84	0,64	0,89	0,77	0,90	0,96	0,70	0,91	0,85	0,83
ACO	-0,26	0,76	0,37	0,72	0,53	0,68	0,68	0,71	0,46	0,70	0,83	0,84	0,86	0,86	0,97	0,91	0,90
CMB	-0,26	0,59	0,35	0,57	0,52	0,66	0,64	0,64	0,45	0,65	0,70	0,70	0,79	0,82	0,82	0,81	0,80
HAD	-0,21	0,67	0,34	0,69	0,69	0,83	0,86	0,77	0,58	0,83	0,84	0,87	0,92	0,74	0,89	0,79	0,83
HAT	-0,23	0,65	0,27	0,67	0,70	0,83	0,86	0,76	0,60	0,83	0,85	0,90	0,92	0,71	0,89	0,81	0,83
REG	-0,29	0,69	0,37	0,70	0,67	0,83	0,83	0,83	0,60	0,87	0,82	0,91	0,91	0,65	0,86	0,84	0,84
PRO	-0,13	0,35	0,23	0,40	0,52	0,64	0,47	0,71	0,48	0,72	0,33	0,65	0,67	0,33	0,51	0,48	0,26
HFR	-0,20	0,84	0,44	0,85	0,62	0,72	0,62	0,65	0,76	0,73	0,73	0,69	0,70	0,43	0,68	0,61	0,78
HEL	-0,13	0,66	0,45	0,67	0,69	0,88	0,73	0,71	0,74	0,86	0,71	0,73	0,74	0,39	0,69	0,46	0,66
LAU	-0,23	0,85	0,42	0,84	0,58	0,73	0,65	0,70	0,67	0,75	0,79	0,78	0,77	0,58	0,79	0,72	0,86
LAC	-0,08	0,57	0,48	0,61	0,71	0,88	0,69	0,70	0,76	0,87	0,60	0,67	0,68	0,21	0,56	0,33	0,48
VID	-0,29	0,85	0,44	0,82	0,57	0,75	0,71	0,75	0,64	0,79	0,85	0,85	0,83	0,66	0,88	0,81	0,89
TLC	-0,24	0,83	0,43	0,83	0,63	0,81	0,74	0,76	0,70	0,83	0,84	0,85	0,85	0,62	0,86	0,77	0,86
TLF	-0,31	0,79	0,41	0,77	0,58	0,76	0,75	0,76	0,62	0,79	0,86	0,86	0,85	0,69	0,90	0,83	0,89
TVC	-0,24	0,76	0,38	0,77	0,65	0,81	0,79	0,78	0,66	0,83	0,82	0,87	0,89	0,70	0,88	0,84	0,86
HOR	-0,25	0,80	0,40	0,79	0,61	0,79	0,75	0,76	0,65	0,81	0,82	0,85	0,88	0,68	0,89	0,82	0,87
INT	-0,24	0,77	0,40	0,79	0,66	0,82	0,74	0,74	0,72	0,84	0,78	0,82	0,85	0,58	0,83	0,74	0,82
CPU	-0,16	0,73	0,40	0,77	0,72	0,88	0,76	0,78	0,75	0,89	0,70	0,83	0,89	0,53	0,81	0,71	0,74

Cuadro I. Matriz de correlaciones (continuación)

	COC	ACO	CMB	HAD	HAT	REG	PRO	HFR	HEL	LAU	VID	TLC	TLF	TVC	HOR	INT	CPU
NRE	-0,26	-0,26	-0,26	-0,21	-0,23	-0,29	-0,13	-0,20	-0,13	-0,23	-0,29	-0,24	-0,31	-0,24	-0,25	-0,24	-0,16
IND	0,73	0,76	0,59	0,67	0,65	0,69	0,35	0,84	0,66	0,85	0,85	0,83	0,79	0,76	0,80	0,77	0,73
PIN	0,39	0,37	0,35	0,34	0,27	0,37	0,23	0,44	0,45	0,42	0,44	0,43	0,41	0,38	0,40	0,40	0,40
DIS	0,73	0,72	0,57	0,69	0,67	0,70	0,40	0,85	0,67	0,84	0,82	0,83	0,77	0,77	0,79	0,79	0,77
TVI	0,74	0,53	0,52	0,69	0,70	0,67	0,52	0,62	0,69	0,58	0,57	0,63	0,58	0,65	0,61	0,66	0,72
PIS	0,87	0,68	0,66	0,83	0,83	0,83	0,64	0,72	0,88	0,73	0,75	0,81	0,76	0,81	0,79	0,82	0,88
PAR	0,83	0,68	0,64	0,86	0,86	0,83	0,47	0,62	0,73	0,65	0,71	0,74	0,75	0,79	0,75	0,74	0,76
REV	0,84	0,71	0,64	0,77	0,76	0,83	0,71	0,65	0,71	0,70	0,75	0,76	0,76	0,78	0,76	0,74	0,78
TCH	0,64	0,46	0,45	0,58	0,60	0,60	0,48	0,76	0,74	0,67	0,64	0,70	0,62	0,66	0,65	0,72	0,75
CIE	0,89	0,70	0,65	0,83	0,83	0,87	0,72	0,73	0,86	0,75	0,79	0,83	0,79	0,83	0,81	0,84	0,89
TAG	0,77	0,83	0,70	0,84	0,85	0,82	0,33	0,73	0,71	0,79	0,85	0,84	0,86	0,82	0,82	0,78	0,70
DAG	0,90	0,84	0,70	0,87	0,90	0,91	0,65	0,69	0,73	0,78	0,85	0,85	0,86	0,87	0,85	0,82	0,83
BAN	0,96	0,86	0,79	0,92	0,92	0,91	0,67	0,70	0,74	0,77	0,83	0,85	0,85	0,89	0,88	0,85	0,89
INO	0,70	0,86	0,82	0,74	0,71	0,65	0,33	0,43	0,39	0,58	0,66	0,62	0,69	0,70	0,68	0,58	0,53
BOT	0,91	0,97	0,82	0,89	0,89	0,86	0,51	0,68	0,69	0,79	0,88	0,86	0,90	0,88	0,89	0,83	0,81
DES	0,85	0,91	0,81	0,79	0,81	0,84	0,48	0,61	0,46	0,72	0,81	0,77	0,83	0,84	0,82	0,74	0,71
EXC	0,83	0,90	0,80	0,83	0,83	0,84	0,26	0,78	0,66	0,86	0,89	0,86	0,89	0,86	0,87	0,82	0,74
COC	1,00	0,92	0,85	0,91	0,92	0,96	0,65	0,81	0,75	0,86	0,92	0,92	0,93	0,95	0,95	0,92	0,93
ACO	0,92	1,00	0,88	0,88	0,87	0,86	0,49	0,73	0,64	0,85	0,91	0,88	0,92	0,91	0,92	0,85	0,81
CMB	0,85	0,88	1,00	0,83	0,82	0,84	0,42	0,66	0,53	0,74	0,80	0,79	0,84	0,85	0,84	0,80	0,76
HAD	0,91	0,88	0,83	1,00	0,99	0,94	0,41	0,75	0,73	0,80	0,88	0,88	0,91	0,92	0,90	0,87	0,82
HAT	0,92	0,87	0,82	0,99	1,00	0,95	0,43	0,77	0,74	0,82	0,89	0,90	0,92	0,93	0,92	0,89	0,85
REG	0,96	0,86	0,84	0,94	0,95	1,00	0,51	0,83	0,73	0,86	0,92	0,93	0,95	0,96	0,95	0,92	0,89
PRO	0,65	0,49	0,42	0,41	0,43	0,51	1,00	0,37	0,52	0,45	0,48	0,50	0,47	0,52	0,51	0,51	0,66
HFR	0,81	0,73	0,66	0,75	0,77	0,83	0,37	1,00	0,78	0,95	0,92	0,94	0,89	0,89	0,90	0,92	0,87
HEL	0,75	0,64	0,53	0,73	0,74	0,73	0,52	0,78	1,00	0,82	0,79	0,84	0,76	0,77	0,79	0,85	0,87
LAU	0,86	0,85	0,74	0,80	0,82	0,86	0,45	0,95	0,82	1,00	0,98	0,98	0,95	0,94	0,96	0,95	0,90
LAC	0,67	0,48	0,40	0,62	0,64	0,64	0,57	0,70	0,95	0,68	0,66	0,73	0,64	0,66	0,66	0,75	0,81
VID	0,92	0,91	0,80	0,88	0,89	0,92	0,48	0,92	0,79	0,98	1,00	0,99	0,99	0,98	0,99	0,96	0,91
TLC	0,92	0,88	0,79	0,88	0,90	0,93	0,50	0,94	0,84	0,98	0,99	1,00	0,98	0,98	0,99	0,98	0,95
TLF	0,93	0,92	0,84	0,91	0,92	0,95	0,47	0,89	0,76	0,95	0,99	0,98	1,00	0,99	0,99	0,96	0,90
TVC	0,95	0,91	0,85	0,92	0,93	0,96	0,52	0,89	0,77	0,94	0,98	0,98	0,99	1,00	0,99	0,97	0,93
HOR	0,95	0,92	0,84	0,90	0,92	0,95	0,51	0,90	0,79	0,96	0,99	0,99	0,99	0,99	1,00	0,98	0,95
INT	0,92	0,85	0,80	0,87	0,89	0,92	0,51	0,92	0,85	0,95	0,96	0,98	0,96	0,97	0,98	1,00	0,97
CPU	0,93	0,81	0,76	0,82	0,85	0,89	0,66	0,87	0,87	0,90	0,91	0,95	0,90	0,93	0,95	0,97	1,00

Cuadro II. Análisis factorial de componentes principales**Total Variance**

Compone	Initial			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of	Cumulative	Total	% of	Cumulative	Total	% of	Cumulative
1	25,5499	72,9998	72,9998	25,550	73,000	73,000	12,636	36,103	36,103
2	2,53690	7,24830	80,2481	2,537	7,248	80,248	8,139	23,255	59,357
3	1,68442	4,81262	85,0607	1,684	4,813	85,061	7,894	22,555	81,912
4	1,15079	3,28796	88,3486	1,151	3,288	88,349	2,053	5,866	87,779
5	1,07326	3,06646	91,4151	1,073	3,066	91,415	1,273	3,637	91,415
6	,76177	2,17650	93,5916						
7	,48988	1,39967	94,9913						
8	,41398	1,18279	96,1740						
9	,36959	1,05598	97,2300						
10	,26735	,76385	97,9939						
11	,19922	,56920	98,5631						
12	,14244	,40697	98,9700						
13	,09484	,27097	99,2410						
14	,07402	,21149	99,4525						
15	,06520	,18629	99,6388						
16	,03869	,11053	99,7493						
17	,03484	,09954	99,8489						
18	,01516	,04330	99,8922						
19	,01211	,03461	99,9268						
20	,01085	,03100	99,9578						
21	,00662	,01892	99,9767						
22	,00427	,01221	99,9889						
23	,00316	,00902	99,9979						
24	,00070	,00201	100,0000						
25	,00000	,00000	100,0000						
26	,00000	,00000	100,0000						
27	,00000	,00000	100,0000						
28	,00000	,00000	100,0000						
29	,00000	,00000	100,0000						
30	,00000	,00000	100,0000						
31	,00000	,00000	100,0000						
32	,00000	,00000	100,0000						
33	,00000	,00000	100,0000						
34	,00000	,00000	100,0000						
35	,00000	,00000	100,0000						

Extraction Method: Principal Component