

EL MONITOREO COMO TÉCNICA DE DETECCIÓN DE LA FALTA DE INFORMACIÓN EN ENCUESTAS SOCIO-ECONÓMICAS

Badler, C.; Alsina, S.; Beltrán, C.; Fracchia, F.; Puigsubirá, C.; Vitelleschi, M.S.

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

INTRODUCCIÓN

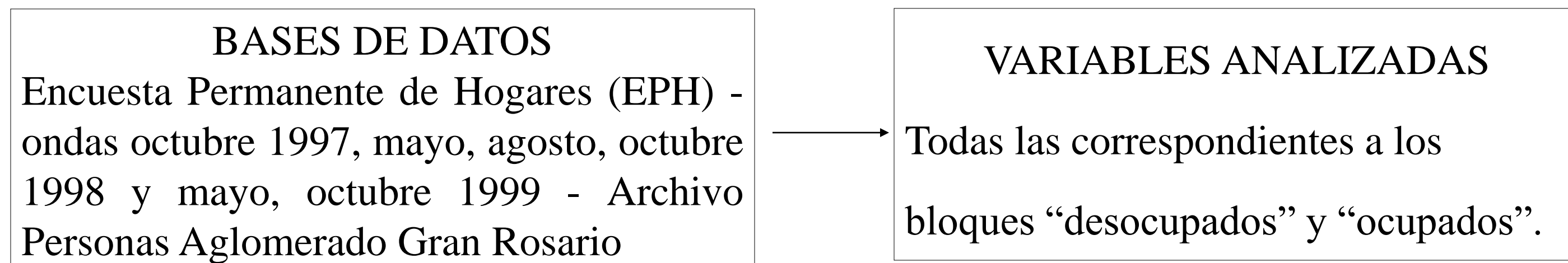
Problemas en la información de bases de datos que afectan y limitan el análisis

Información faltante: existen dificultades en su captación e identificación.

Seguimiento de casos a través de monitoreo para establecer la real presencia de información faltante.

MATERIAL

El material utilizado es heterogéneo, debido a los objetivos del trabajo



Cuestionarios y manuales de procedimientos y contenidos de la EPH

Softwares: SAS 8.1 , Microsoft Excel 2000

METODOLOGÍA

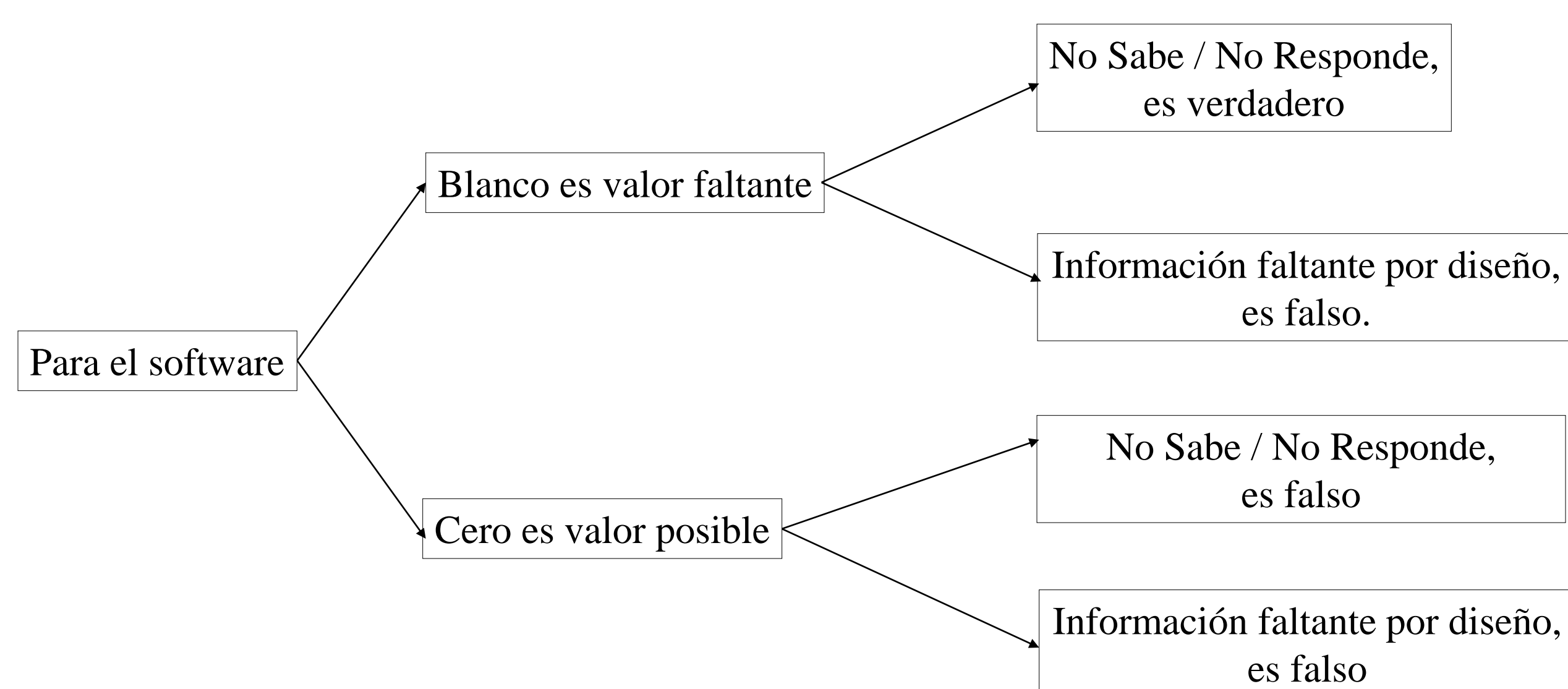
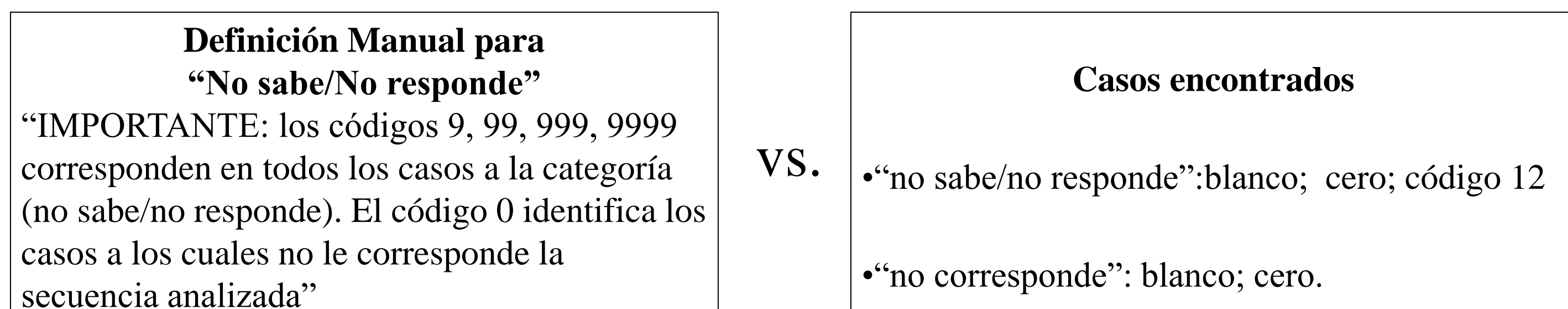
MONITOREO

- Importación de la base de datos (formato, software)
- Estructura de la base (características de la variable)
- Distribuciones de frecuencias
- Identificación de la información faltante (códigos, blancos)
- Seguimientos de casos con información faltante (pérdidas parciales y totales)
- Pruebas de consistencia (diferenciar información perdida, confusa y faltante por diseño)
- Descripción de patrones de asociación y detección de dependencias estadísticas: se utiliza el modelo log-lineal.

RESULTADOS

MONITOREO

- Los softwares no presentan diferencias al levantar, captar e identificar los datos. Se verifica:
- Presencia y características de las variables de la base estudiada (Proc CONTENTS).
- Valores observados, categorías y códigos (Proc FREQ).
- Presencia de inconsistencias, información faltante e información confusa (como la que no corresponde a las categorías asignadas).
- Distintas formas de identificación de la información faltante.



COMPARACION BLOQUES DESOCUPADOS Y OCUPADOS A PARTIR DEL MONITOREO

- Variables exclusivas bloque desocupados: mayor frecuencia de valores perdidos en variables de tiempo transcurrido desde ocupación anterior y características del establecimiento.
- Variables comunes a los dos bloques: mayor proporción de información faltante en variables relacionadas al ingreso, migraciones y características ocupacionales.
- 19 Variables comunes a los dos bloques tienen pérdida (Tablas 1 y 2).

DISCUSIÓN

- La técnica de Monitoreo debe ser aplicada previamente a la descripción y análisis de bases de datos provenientes de encuestas y registros.
- Conjuntamente con la evaluación de la calidad de la información esta técnica permite detectar la información faltante y/o confusa que no siempre es directamente visualizable. Para su implementación se requiere la consideración simultánea de manuales de procedimientos, técnicas estadísticas y elementos computacionales.
- La información faltante y/o confusa puede surgir en cada una de las etapas necesarias para la realización de encuestas y registros y obedecen a problemas multicausales, pero al detectarse pueden aplicarse tratamientos para mejorar la calidad de los datos a analizar.

Tabla 1. Número de variables con falta de información según onda y bloque.

Bloque	Onda											
	Oct-97		May-98		Ago-98		Oct-98		May-99		Oct-99	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
Desocupados	7	37	6	32	3	16	5	26	5	26	6	32
Ocupados	12	63	9	47	11	58	13	68	11	58	10	53

Tabla 2. Cantidad de variables con falta de información según porcentaje y unidades afectadas, bloque y onda.

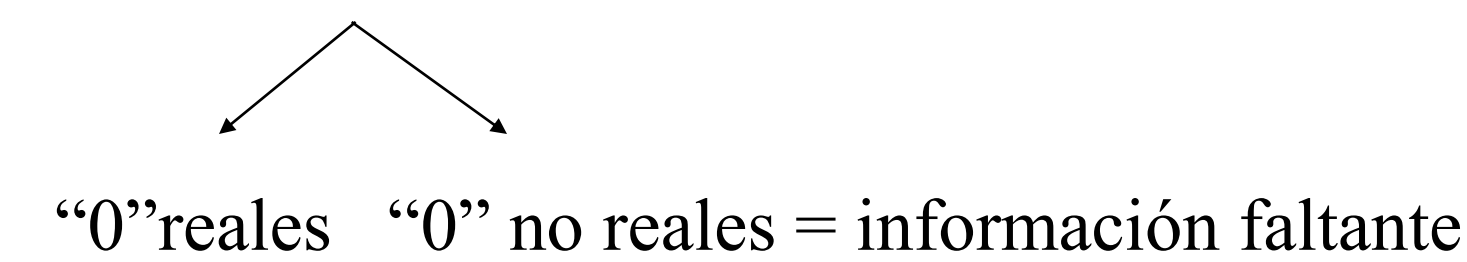
Porcentaje de variables	Onda											
	Oct-97		May-98		Ago-98		Oct-98		May-99		Oct-99	
	Des.	Ocu.	Des.	Ocu.	Des.	Ocu.	Des.	Ocu.	Des.	Ocu.	Des.	Ocu.
(0.01 - 1]	5	5	3	2	0	3	4	6	1	5	0	5
(1 - 2]	1	2	2	2	1	3	0	2	3	1	4	0
(2 - 5]	0	0	1	0	2	0	0	0	0	0	1	0
(5 y +)	1	5	0	5	0	5	1	5	1	5	1	5

Tabla 3. Cantidad de unidades con valor cero verificadas como información faltante según variables de ingresos, bloque y onda.

Variable	Onda											
	Oct-97		May-98		Ago-98		Oct-98		May-99		Oct-99	
	Desoc.	Ocup.	Desoc.	Ocup.	Desoc.	Ocup.	Desoc.	Ocup.	Desoc.	Ocup.	Desoc.	Ocup.
	n=206	n=1299	n=214	n=1260	n=162	n=1177	n=170	n=1033	n=159	n=873	n=176	n=856
P47T	163 (1)	124 (89)	163 (1)	99 (69)	120 (0)	102 (79)	129 (0)	108 (92)	115 (3)	109 (98)	140 (2)	107 (90)
ITF-IPCF	33 (12)	126 (122)	22 (7)	96 (94)	21 (7)	124 (123)	40 (15)	126 (125)	22 (19)	144 (144)	37 (18)	124 (124)

VARIABLES RELACIONADAS AL INGRESO: P47T, ITF e IPCF

- Sin blancos ni códigos especiales de falta de información
- Con categoría "0"



RELACIÓN ENTRE ITF=0, ESTADO OCUPACIONAL Y ONDA

Tabla 4. Cantidad de personas clasificadas según estado, onda e ingreso total familiar.

Estado	Onda						Total
	Oct 97		Ago 98		Oct 99		
	ITF = 0	ITF > 0	ITF = 0	ITF > 0	ITF = 0	ITF > 0	
Ocupado	64	652	69	605	64	405	1859
Desocupado	29	144	18	116	30	117	454
Total	93	796	87	721	94	522	2313

El modelo log-lineal resultante contiene las asociaciones ITF*Onda, Onda*Estado, ITF*Estado ($G^2=1.26$; $p=0,5326$; $g.l.=2$).

Tabla 5. Razones de odds estimadas del modelo e intervalos de confianza del 95%

	Estimación de las razones de odds	Intervalo de confianza del 95%
ITF*ESTADO	0.595	(0.4465;0.7936)
ITF*ONDA		
Onda 97 vs. 98	0.9520	(0.6980;1.2980)
Onda 97 vs. 99	0.6638	(0.4876;0.9036)
ONDA*ESTADO		
Onda 97 vs. 98	0.8205	(0.6394;1.053)
Onda 98 vs. 99	1.5406	(1.0470;2.266)

En las tres ondas, los ingresos nulos se presentaron principalmente en aquellos hogares donde habitan desocupados. Asimismo, en los hogares encuestados en la onda de octubre de 1999 se observaron ingresos familiares nulos más frecuentemente que en las ondas de agosto de 1998 y octubre de 1997.