

**EGM 2008
UN AÑO DE FUSION
REFLEXIONES Y EXPERIENCIAS**



Madrid, 4 Febrero 2009

PUNTO de PARTIDA

Coexistencia de fuentes diversas.

Desde el punto de vista del objeto de estudio:

- Fuente Multimedia (EGM)
- Fuentes Monomedia

Desde el punto de vista del objeto de estudio:

- Entrevistas personales en el hogar, muestreo aleatorio estratificado polietápico por rutas aleatorias.
- Entrevistas telefónicas por marcación aleatoria de números telefónicos

Punto de partida: Coexistencia de diferentes fuentes

EGM 3ª OLA 2008		Entrevistas:							
MULTIMEDIA	10.285	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS Internet Cine Exterior	
MONOMEDIA PRENSA Telefónico	15.089	DEMO GRAFICOS		PRENSA					
MONOMEDIA RADIO Telefónico	16.524	DEMO GRAFICOS			RADIO				
MONOMEDIA REVISTAS Personal	6.789	DEMO GRAFICOS				REVISTAS			
MONOMEDIA TV Telefónico	4.315	DEMO GRAFICOS					TV		



PUNTO de PARTIDA

Coexistencia de fuentes diversas.

Hasta 5 fuentes, ficheros, distintos, que suministran cada uno un dato "oficial".

Inconvenientes:

- Dato de audiencia diferente para los medios con ampliación siempre que se necesita cruzar con otros medios u obtener datos de audiencia cruzados con equipamiento, consumo, etc.
- Bases diferentes para EGM Multimedia y cada uno de los medios con ampliación.

EGM RADIO Dato Oficial: Audiencia Radio Entrevistas: 26.809	MULTIMEDIA 10.285 Personal
	+ MONOMEDIA RADIO 16.524 Telefónico

EGM PRENSA Dato Oficial: Audiencia Prensa Entrevistas: 25.374	MULTIMEDIA 10.285 Personal
	+ MONOMEDIA PRENSA 15.089 Telefónico

EGM MULTIMEDIA Dato Oficial: Estilos de Vida Equipamiento Consumos Sociodemográficos Otros Medios	MULTIMEDIA 10.285 Personal
--	--






EGM REVISTAS Dato Oficial: Audiencia Revistas Entrevistas: 17.074	MULTIMEDIA 10.285 Personal
	+ MONOMEDIA REVISTAS 6.789 Personal

EGM TV Dato Oficial: Audiencia TV Entrevistas: 14.600	MULTIMEDIA 10.285 Personal
	+ MONOMEDIA TV 4.315 Telefónico

FUENTES OFICIALES

OBJETIVO






OBJETIVO:
 Obtener un solo fichero con la información de todos los medios y con información en todos los ítems “no medios”

EGM 3ª OLA 2008	Entrevistas:								
MULTIMEDIA  Personal	10.285	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS	
MONOMEDIA PRENSA  Telefónico	15.089	DEMO GRAFICOS		PRENSA					
MONOMEDIA RADIO  Telefónico	16.524	DEMO GRAFICOS			RADIO				
MONOMEDIA REVISTAS  Personal	6.789	DEMO GRAFICOS				REVISTAS			
MONOMEDIA TV  Telefónico	4.315	DEMO GRAFICOS					TV		
TOTAL	53.002	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS	

OBJETIVO: Fuente Única

OBJETIVO

- El problema planteado es un problema de imputación: “RELLENAR LOS HUECOS” o “FALTAS DE INFORMACION”.
 - Sujeto a ciertas restricciones:
- Premisa fundamental:**
respetar al 100% el dato de audiencia oficial de cada medio.
- Premisas accesorias:**
Método trazable, transparente, Automatización máxima y no debe suponer un retraso en la publicación de las olas.

		FUSION							
		INFOR MACION COMUN	INFOR MACION IMPUTADA						
EGM 3ª OLA 2008									
MULTIMEDIA  Personal	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS		
	+	+	+	+	+	+	+	+	+
MONOMEDIA PRENSA  Telefónico	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS		
	+	+	+	+	+	+	+	+	+
MONOMEDIA RADIO  Telefónico	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS		
	+	+	+	+	+	+	+	+	+
MONOMEDIA REVISTAS  Personal	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS		
	+	+	+	+	+	+	+	+	+
MONOMEDIA TV  Telefónico	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS		
	+	+	+	+	+	+	+	+	+
EGM FUSION	DEMO GRAFICOS	ESTILOS VIDA EQUIPAMIENTO CONSUMOS	PRENSA	RADIO	REVISTAS	TV	OTROS MEDIOS		
	=	=	=	=	=	=	=	=	=



“Los datos del fichero resultante deben ser los mismos que los de la fuente de la que proceden, que el dato oficial”

¿Qué quiere decir esto?

Uno de los objetivos perseguidos es sacar mayor partido de la información disponible.

¿Para que se hacen “ampliaciones”?

- ¿Para obtener un valor de la audiencia total distinto?
- ¿Para obtener mejores estimaciones cuando vayamos profundizando en el análisis, cuando vayamos desagregando?

Penetración:

Variables de control con distinta distribución:

Supongamos una revista sólo para universitarios que tiene una penetración del 20% en el fichero multimedia.

	MULTIMEDIA	MONOMEDIA RADIO	
		Estimación basada en penetración	Estimación basada en lectores
Universitarios	1.600.000	1.200.000	1.200.000
Penetración Revista	20%	20%	27%
Lectores	320.000	240.000	320.000

¿Es posible obtener los mismos valores de audiencia cuando los ficheros son diferentes?

COMPONENTES DEL PROBLEMA

GENERALES O TEÓRICAS:

Procedimientos generales para sustituir los valores que faltan.

Normalmente consideran una única población de individuos equivalentes, pocas variables, independientes o casi.

PARTICULARES O IDIOSINCRÁTICAS:

- Muestreo.
- Ponderación.
- Equilibraje
- Variables locales.
- Variables temporales
- Trazabilidad.
- Coexistencia de universos: individuos y hogares.

COMPONENTES TEORICAS

Soluciones posibles:

- Fusión pura
- Estimación

La fusión pura conserva la estructura, correlaciones, de las variables imputadas y es más adecuada cuando se cuenta con poca información para inferir un modelo.

La aproximación por cesión o fusión pura permite trabajar directamente con la codificación condensada, la grabación directa del cuestionario, mientras que la aproximación por estimación exige la elaboración de las variables reales, (oye/n oye), es decir el tratamiento integro de la información antes de proceder a la fusión.

FUSION

CESION

Donante:

AAAXXXX
BBBYYYY
.....

Fusionado:

AABXXXX
BABYYYY
.....

Receptor:

AAB ____
BAB ____
.....

Técnicas:

- Random hot deck
- Distancias entre individuos (KNN)
- Fusión sobre Referencial Factorial
- Segmentación

ESTIMACION

Donante:

AAAXXXX
BBBYYYY
.....

MODELO

Fusionado:

AABYXXY
BABYXYX
.....

Receptor:

AAB ____
BAB ____
.....

Técnicas:

- Análisis Discriminante
- Regresiones logísticas
- Redes Neuronales Artificiales
- Métodos Bayesianos / Redes

COMPONENTES DE LA SOLUCION ADOPTADA

1. PONDERACIÓN COMÚN: EQUILIBRAJE Y ELEVACION
2. ESTRATIFICACIÓN COMÚN
3. TRANSFERENCIA / TRANSPLANTE DE INFORMACIÓN:
ASIGNACIÓN MÚLTIPLE



TAMAÑO 14/O MAS DEL HOGAR	PROVINCIA	SEXO	EDAD	DIA DE LA ENTREVISTA	DIA DE LA SEMANA
3	50	2	7	7	2

1	Alava	La Rioja	Hombre	14 a 19	Lunes	Lunes - Viernes
2	Albacete	Lugo	Mujer	20 a 24	Martes	Sábado - Domingo
3 / MAS	Alicante	Madrid		25 a 34	Miércoles	
	Almería	Málaga		35 a 44	Jueves	
	Avila	Murcia		45 a 54	Viernes	
	Badajoz	Navarra		55 a 64	Sábado	
	Baleares	Orense		65 y más	Domingo	
	Barcelona	Oviedo				
	Burgos	Palencia				
	Cáceres	Las Palmas				
	Cádiz	Pontevedra				
	Castellón	Salamanca				
	Ciudad Real	S.C.Tenerife				
	Córdoba	Santander				
	Coruña La	Segovia				
	Cuenca	Sevilla				
	Gerona	Soria				
	Granada	Tarragona				
	Guadalajara	Teruel				
	Guipúzcoa	Toledo				
	Huelva	Valencia				
	Huesca	Valladolid				
	Jaén	Vizcaya				
	León	Zamora				
	Lérida	Zaragoza				

VARIABLES DE EQUILIBRAJE

HABITAT	HABITAT	AMA DE CASA	CABEZA DE FAMILIA	REGION	REGION
2	7	2	2	17	9

Menos de 50.000	Hasta 2000	Si	Si	Andalucía	Andalucía
Mas de 50.000 + Capitales	De 2 a 5.	No	No	Aragón	Castilla/León
	De 5 a 10.			Asturias	Catalana
	De 10 a 50.			Baleares	País Valenciano
	De 50. a 200.			Canarias	Galicia
	De 200 a 500.			Cantabria	Madrid
	Capitales			Castilla/León,	País Vasco
				Castilla/Mancha	Uniprovincial.
				Catalana	Resto
				C.Valenciana	
				Extremadura	
				Galicia	
				Madrid	
				Murcia	
				Navarra	
				País Vasco	
				La Rioja	

EQUILIBRAJE:

MATRICES DE EQUILIBRAJE TRADICIONALES (Individuos)

Matriz: Variables (Número de categorías):

1	PROVINCIA (50)	X	HABITAT (2)
2	REGION (17)	X	HABITAT (7)
3	REGION (17)	X	SEXO (2)
4	REGION (17)	X	EDAD (7)
5	EDAD (7)	X	SEXO (2)
6*	DIA SEMANA (7)	X	AMA DE CASA (2)
7	CABEZA FAMILIA (2)		
8	REGION (9)	X	TAMAÑO HOGAR (3)
9	REGION (9)	X	DIA SEMANA (2)

* En el caso de Revistas esta matriz se sustituye por AMA DE CASA (2)



EQUILIBRAJE:

MATRICES DE EQUILIBRAJE NUEVAS (Individuos)






Matriz: Variables (Número de categorías):

10	PROVINCIA (50)	X	SEXO (2)	X	HABITAT (2)
11	PROVINCIA (50)	X	DIA SEMANA (2)	X	HABITAT (2)
12	PROVINCIA (50)	X	DIA SEMANA (7)		
13	HABITAT (2)	X	DIA SEMANA (7)		
14	HABITAT (7)	X	DIA SEMANA (7)		
15	REGION (17)	X	DIA SEMANA (7)		
16	PROVINCIA (50)		DIA SEMANA (2)		
17	HABITAT (2)	X	DIA SEMANA (2)		
18	HABITAT (7)	X	DIA SEMANA (2)		
19	REGION (17)	X	DIA SEMANA (2)		

* Para el equilibraje del fichero de revistas se suprimen las matrices que contienen el día de la semana



ESTRATIFICACION:

Paso	Información	Donante	Receptor	Estratificación
1	Estilos vida Consumos individuo Otras audiencias	EGM Multimedia	Resto	Provincia (50) x Hábitat (2) x Sexo (2) 
2	Consumos hogar	Amas de Casa EGM Multimedia	Amas casa Resto	Día semana (7) 
3	Audiencia Radio	EGM Radio (Multimedia+ Telefónico Radio)	Resto	Provincia (50) x Hábitat (2) x Día semana (2) 
4	Audiencia Prensa	EGM Prensa (Multimedia+ Telefónico Prensa)	Resto	Provincia (50) x Hábitat (2) x Día Semana (2) 
5	Audiencia Revistas	EGM Revistas (Multimedia+ Personal Revistas)	Resto	Provincia (50) x Hábitat (2) x Sexo (2) 
6	Audiencia TV	EGM TV (Multimedia+ Telefónico TV)	Resto	Provincia (50) x Hábitat (2) x Día semana (2) 

SIMILARIDAD O DISTANCIA ENTRE INDIVIDUOS:

VARIABLES UTILIZADAS EN EL CÁLCULO DE LA DISTANCIA O SIMILARIDAD ENTRE INDIVIDUOS DENTRO DEL MISMO ESTRATO:

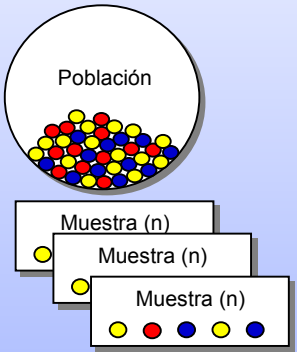
CLASE SOCIAL
TAMAÑO DEL HOGAR
PROVINCIA
SEXO
EDAD
DIA DE LA ENTREVISTA
DIA DE LA SEMANA
HABITAT
HABITAT
AMA DE CASA
CABEZA DE FAMILIA
REGION
REGION
CONOCIMIENTO IDIOMA AUTONOMICO
USO IDIOMA AUTONOMICO
MUNICIPIO

Metodología:

La aportación de las variables nominales a la distancia se calcula sumando 1 a la distancia total cuando la categorías a que pertenece cada individuo no coinciden. En el caso de variables ordinales como la clase social o la edad se calcula la diferencia entre los valores de los dos individuos, tomando el valor absoluto de la diferencia y dividiéndolo por 2.

Una vez calculadas todas las distancias entre donantes y receptores se da una primera vuelta produciéndose solo el trasvase de información si los dos individuos residen en el mismo municipio, y una segunda vuelta en que se permite el trasvase de la información restante aunque los individuos no residan en el mismo municipio.

EGM Multimedia y Monomedias



$$\sum w_d = \sum w_r$$

- 1) Calculamos la matriz de distancias entre donantes y receptores utilizando la información tanto de las variables originales de control como del resto de variables, pero en esta ocasión no hemos replicado ni donantes ni receptores, sólo tenemos los individuos originalmente en los ficheros reales.
- 2) Ordenamos las distancias de menor a mayor y seleccionamos aquellos que son más parecidos.

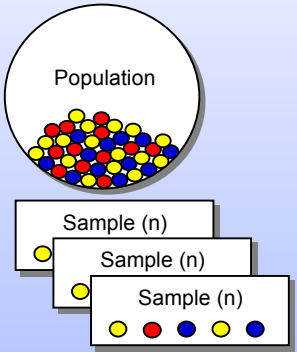
Para un estrato h tendremos:

		Receptores						
		1	2	j	...	q
Peso		wr1	wr2	wrj	...	wrq
Donantes	Peso							
	1	wd1						
	2	wd2						
						
	i	wdi						
						
	p	wdp						

Matriz de distancias

- 3) Si el peso del donante wdi es mayor que el del receptor, imputamos al receptor la información del donante, lo eliminamos de la tabla de candidatos y pasa al fichero de receptores con su peso. El donante permanece como tal donante en la tabla pero con un peso igual a la diferencia entre su peso y el del receptor al que acaba de transmitir su información.
- 4) Si el peso del receptor es mayor, le trasmitimos la información del donante, y lo pasamos al fichero definitivo con un peso igual al del donante. Borramos el donante de la tabla de donante y dejamos al receptor con un peso igual al de la diferencia de pesos.
- 5) Así hasta que desaparezcan donantes y receptores.

FUSION



- 1) Se calcula la matriz de distancias y se ordenan estas de menor a mayor.
- 2) El par correspondiente a la menor distancia es seleccionado.

DONANTES Y RECEPTORES TIENEN PESOS DIFERENTES

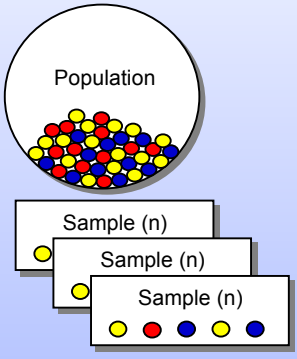
Donantes y sus pesos:	Receptores y sus pesos:					Peso Total receptores:
	3,10	0,60	1,40	1,80	1,10	8,00
3,00						
1,50		8				
2,00	2	6	3			
1,00		1	7			
0,50			5			
8,00						
Peso total						

Matriz de distancias



DONANTES Y RECEPTORES TIENEN DIFERENTES PESOS

FUSION



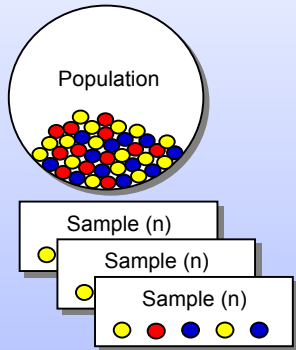
Donantes y sus pesos:	Receptores y sus pesos:					Peso Total
	3,10	0,60	1,40	1,80	1,10	8,00
3,00		▲				
1,50		8				
2,00	2	6	3			
1,00	←	1	7			
0,50		▲	5			
8,00						
Peso total						

Matriz de distancias

- Par más similar: Peso del donante mayor que el del receptor**
- 1) Pegamos la información del donante al receptor
 - 2) El receptor es escrito en el fichero fusionado con su peso y borrado de la tabla.
 - 3) El donante permanece en la tabla con un peso igual a su peso original menos el de el donante al que acaba de transferir su información.

DONANTES Y RECEPTORES TIENEN DIFERENTES PESOS

FUSION



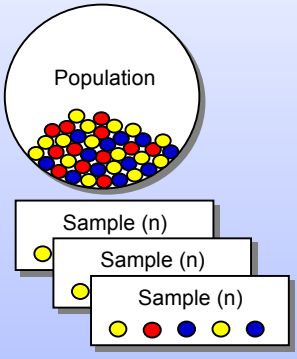
Donantes y sus pesos:	Receptores y sus pesos:					Peso Total
	Receptores:					
	3,10	0,50	1,40	1,80	1,10	7,40
3,00						
1,50		8				
2,00	2	6	3			
0,40		1	7			
0,50			5			
7,40						

Matriz de distancias

- Par más similar: Peso del donante mayor que el del receptor**
- 1) Pegamos la información del donante al receptor
 - 2) El receptor es escrito en el fichero fusionado con su peso y borrado de la tabla.
 - 3) El donante permanece en la tabla con un peso igual a su peso original menos el de el donante al que acaba de transferir su información.

DONANTES Y RECEPTORES TIENEN DIFERENTES PESOS

FUSION



- 1) Se calcula la matriz de distancias y se ordenan estas de menor a mayor.
- 2) El par correspondiente a la menor distancia es seleccionado.

Donantes y sus pesos:	Receptores y sus pesos:					Peso Total
	3,10	0,00	1,40	1,80	1,10	7,40
3,00	↑	8				
1,50	↑	8				
2,00	← 2	8	3			
0,40	↑	1	7			
0,50	↑	5				
7,40						
Peso total						

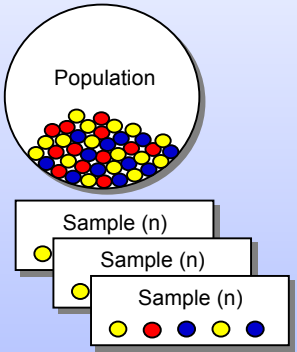
Matriz de distancias

Par más similar: Peso del receptor es mayor que el peso del donante.

- 1) Pegamos la información del donante al receptor.
- 2) El receptor es escrito en el fichero fusionado con un peso igual al del donante y permanece en la tabla de receptores con un peso igual al suyo original menos el del donante del que acaba de recibir la información.
- 3) El donante es borrado de la tabla de donantes: ya ha transferido su información.

DONORS and RECEIVERS HAVE DIFFERENT WEIGHTS

FUSION



El proceso continúa hasta que no queda ningún donante con información que transferir ni ningún receptor a la espera de información.

Donantes y sus pesos:	Receptores y sus pesos:					Peso Total
	Receptores:					
	1,10	0,00	1,40	1,80	1,10	5,40
3,00	↑	8				
1,50	↑	6				
0,00	2	4	3			
0,40	↑	1	7			
0,50	↑		5			
5,40						
Peso total						

Matriz de distancias

- Par más similar: Peso del receptor es mayor que el peso del donante.
- 1) Pegamos la información del donante al receptor.
 - 2) El receptor es escrito en el fichero fusionado con un peso igual al del donante y permanece en la tabla de receptores con un peso igual al suyo original menos el del donante del que acaba de recibir la información.
 - 3) El donante es borrado de la tabla de donantes: ya ha transferido su información.

FUSION

RESULTADO:

El fichero donante, el receptor y la suma de ambos contienen exactamente la **misma información para las variables a imputar** en todos los estratos de la post estratificación común y en todas sus combinaciones.

Además las **relaciones internas**, correlaciones, de todas las variables imputadas son las mismas en todos los ficheros.

En las variables no controladas las distribuciones son lo más parecidas posibles.

Se cumple el requerimiento de **trazabilidad**. Ej: si en donante, multimedia de 14.000 entrevistas hay 7.000 entrevistas que declaran leer el soporte X, y que ponderadas suponen 9.000 lectores, en el receptor de 10.000 entrevistas habrá, 5.000 entrevistas, calculadas mediante la ponderación muestral, que contengan esa información y que ponderadas supondrán 9.000 lectores.

A mayor número de variables controladas mayor igualdad en las distribuciones pero estratos más pequeños y menor posibilidad de encontrar individuos parecidos. Hay que buscar un **equilibrio entre número y tamaño de estratos**.

AUTOMATIZACION

The screenshot shows the FUSION - DIGITAL Visual Fortran IDE. The main window displays a Fortran program with the following code:

```

I2=0
I3=0
AP2 = ' 10000'
1 READ(10,901,END=999)A
901 FORMAT(A14242)
READ(A(159:159),'(I1)')ITIPO
IF(J.NE.ITIPO) GOTO 1
IF (ITIPO.EQ.1) THEN
  I1 = I1 + 1
  READ(11,911)AP1
  A(6739:6746) = AP1
  A(6757:6764) = AP2
  READ(13,931)WHO(I1)
  WRITE(A(6766:6773),'(I8)')INT(WHO(I1)*10000.)
  AA(I1) = A
  WRITE(1,901)A
ELSE IF (ITIPO.EQ.2) THEN
  I2 = I2 + 1
  READ(11,951)WIO(I2)
  READ(13,931)WHO(I2)
  AA(I2) = A
ELSE IF (ITIPO.EQ.3) THEN
  I3 = I3 + 1
  READ(11,951)WIO(I3)
  READ(13,931)WHO(I3)
  AA(I3) = A
END IF
GOTO 1
999 CONTINUE
911 FORMAT(5X,16X,3(8X),A8)
951 FORMAT(5X,16X,3(8X),F8.4)
931 FORMAT(16X3(8X),F8.4)

```

The left pane shows a project tree for 'Workspace FUSION: 6 project(s)'. The tree includes folders for 'FUSION files', 'FUSIONA files', 'FUSIONA_2AFASE_1 files', 'FUSIONA_2AFASE_2 files', 'PONDERA_H files', and 'SUTURA files'. Each folder contains sub-folders for 'Source Files', 'Header Files', and 'Resource Files'. The 'SUTURA files' folder is currently selected, showing 'SUTURA.FOR' as the active file.

RESULTADOS
DE LA FUSION

Tamaño y Peso relativo de
las fuentes

EGM: 3ª OLA 2008

	Entrevistas en cada fuente	Registros en el fichero resultante de la fusión	Registros fichero fusionado / Entrevistas	Entrevistas de cada fuente (\$%)	Peso de cada fuente en el fichero final (%)
Total	53.002	153.249	2,9	100,0	100,0
EGM Personal multimedia	10.285	11.136	1,1	19,4	20,0
Telefónico RADIO	16.524	49.469	3,0	31,2	29,6
Telefónico PRENSA	15.089	49.367	3,3	28,5	28,2
Personal REVISTAS	6.789	26.373	3,9	12,8	13,7
Telefónico TV	4.315	16.904	3,9	8,1	8,5



RESULTADOS DE LA FUSION

Equivalencia cifras globales de audiencia

EGM: 3ª OLA 2008

	FUSIÓN	EGM Multimedia	EGM Radio	EGM Prensa	EGM Revistas
Total	38.261	38.261	38.261	38.261	38.261
Diarios(total)	15.912	15.127		15.912	
Supl. Dominicales	7.811	7.830			
Revistas Semanales	9.488	8.258			9.488
Revistas Quincenales	1.371	1.170			1.371
Revistas Mensuales	16.074	14.435			16.074
Revistas(total)	20.062	18.338			20.062
Radio OM	1.588	1.138	1.588		
Radio FM	19.320	19.539	19.320		
Radio Generalista	10.803	10.307	10.803		
Radio Temática	11.330	11.492	11.330		
Radio(total)	20.865	20.607	20.865		
Cine Últ.Semana	1.341	1.344			
Internet Ayer	12.050	12.054			
Diarios en Internet Ayer	3.093	3.094			
Revistas en Internet Ayer	284	284			
TV en Internet Ayer	423	418			
Exterior Ayer	20.005	20.030			



RESULTADOS DE LA FUSION

Equivalencia cifras de audiencia por estratos

EGM: 3ª OLA 2008

	FUSIÓN	EGM Multimedia	EGM Radio	EGM Prensa	EGM Revistas	EGM Televisión
Individuos > 14 años : MALAGA	1.252	1.250	1.252	1.252	1.252	1.252
Diarios(total)	403	384		403		
Supl. Dominicales	84	84				
Revistas Semanales	192	155			192	
Revistas Quincenales	32	52			32	
Revistas Mensuales	267	290			267	
Revistas(total)	397	384			397	
Radio OM	27	33	27			
Radio FM	638	595	638			
Radio Generalista	283	272	283			
Radio Temática	426	365	426			
Radio(total)	668	613	668			
Total Autonómicas	407	449				410
Aut. Públicas	407	449				410
Aut. Privadas	0	0				0
Aut. Internacio	0	0				0
Temáticas	128	103				125
Temáticas Pago	76	52				73
Resto Temáticas	62	51				62
TV Locales	22	18				23
Resto Cadenas	48	0				48
Total Nacionales	972	988				968
TOTAL TV	1.111	1.135				1.111
Cine Últ.Semana	26	26				
Internet Ayer	338	338				
Diarios en Internet Ayer	92	92				
Revistas en Internet Ayer	4	4				
TV en Internet Ayer	15	15				
Exterior Ayer	603	603				

RESULTADOS DE LA FUSION

Equivalencia correlación variables tratadas conjuntamente a nivel global y por estrato:
Duplicación soportes mismo medio

EGM FUSIÓN 3ª Ola 2008

	Total	LECTORES ULTIMO PERIODO						REVISITAS MENSUALES
		REVISITAS SEMANALES	REVISITAS QUINCENALES	Super Pop	Bravo por ti	Computer Hoy	Pc Today	
LECTORES ULTIMO PERIODO	38.261	9.488	1.371	661	596	450	245	16.074
REVISITAS SEMANALES	9.488	9.488	638	377	311	149	88	5.679
REVISITAS QUINCENALES	1.371	638	1.371	661	596	450	245	1.091
Super Pop	661	377	661	661	410	12	13	500
Bravo por ti	596	311	596	410	596	14	13	437
Computer Hoy	450	149	450	12	14	450	140	407
Pc Today	245	88	245	13	13	140	245	221
REVISTA MENSUALES	16.074	5.679	1.091	500	437	407	221	16.074

EGM REVISTAS 3ª Ola 2008

	Total	LECTORES ULTIMO PERIODO						REVISITAS MENSUALES
		REVISITAS SEMANALES	REVISITAS QUINCENALES	Super Pop	Bravo por ti	Computer Hoy	Pc Today	
LECTORES ULTIMO PERIODO	38.261	9.488	1.371	661	596	450	245	16.074
REVISITAS SEMANALES	9.488	9.488	638	377	311	149	88	5.679
REVISITAS QUINCENALES	1.371	638	1.371	661	596	450	245	1.091
Super Pop	661	377	661	661	410	12	13	500
Bravo por ti	596	311	596	410	596	14	13	437
Computer Hoy	450	149	450	12	14	450	140	407
Pc Today	245	88	245	13	13	140	245	221
REVISTA MENSUALES	16.074	5.679	1.091	500	437	407	221	16.074

RESULTADOS DE LA FUSION

Equipamiento Telefónico

EGM: 3ª OLA 2008 (individuos)

	FUSIÓN	EGM Multimedia	EGM Radio	EGM Prensa	EGM Revistas	EGM Televisión
Población (000)	38.261	38.261	38.261	38.261	38.261	38.261
Ni movil ni fijo	199	282	121	119	310	201
Solo móvil	6.098	9.294	4.755	7.189	9.111	7.877
Solo fijo	2.824	2.682	3.221	3.283	2.766	2.764
Movil y Fijo	29.140	26.003	30.164	27.670	26.074	27.419
Total móvil	35.238	35.297	34.919	34.859	35.185	35.296
Total fijo	31.964	28.685	33.385	30.953	28.840	30.183
Total móvil o fijo	38.062	37.979	38.140	38.142	37.951	38.060
Población (%)	100,0	100,0	100,0	100,0	100,0	100,0
Ni movil ni fijo	0,5	0,7	0,3	0,3	0,8	0,5
Solo móvil	15,9	24,3	12,4	18,8	23,8	20,6
Solo fijo	7,4	7,0	8,4	8,6	7,2	7,2
Movil y Fijo	76,2	68,0	78,8	72,3	68,1	71,7
Total móvil	92,1	92,3	91,3	91,1	92,0	92,3
Total fijo	83,5	75,0	87,3	80,9	75,4	78,9
Total móvil o fijo	99,5	99,3	99,7	99,7	99,2	99,5

Estudios telefónicos introducen un sesgo

Equipamiento Telefónico

RESULTADOS DE LA FUSION

Vías de corrección:

- Imputación equipamiento telefónico.
- Ponderación con base a los datos del Multimedia Personal
- Corrección del sesgo en las fuentes con mayor número de entrevistas telefónicas a hogares sólo móviles.

EGM: 3ª OLA 2008 (hogares)

	EGM				EGM	EGM
	FUSIÓN	Multimedia	EGM Radio	EGM Prensa	Revistas	Televisión
Población (000)	16.122	16.122	16.122	16.122	16.122	16.122
Ni movil ni fijo	118	243	97	93	242	164
Solo móvil	2.631	4.039	2.075	3.125	4.196	3.532
Solo fijo	1.651	1.878	2.170	2.134	1.733	1.757
Movil y Fijo	11.722	9.962	11.780	10.770	9.951	10.669
Total móvil	14.353	14.001	13.855	13.895	14.147	14.201
Total fijo	13.373	11.840	13.950	12.904	11.684	12.426
Total móvil o fijo	16.004	15.879	16.025	16.029	15.880	15.958
Población (%)	100,0	100,0	100,0	100,0	100,0	100,0
Ni movil ni fijo	0,7	1,5	0,6	0,6	1,5	1,0
Solo móvil	16,3	25,1	12,9	19,4	26,0	21,9
Solo fijo	10,2	11,6	13,5	13,2	10,7	10,9
Movil y Fijo	72,7	61,8	73,1	66,8	61,7	66,2
Total móvil	89,0	86,8	85,9	86,2	87,7	88,1
Total fijo	82,9	73,4	86,5	80,0	72,5	77,1
Total móvil o fijo	99,3	98,5	99,4	99,4	98,5	99,0

Nacionalidad

RESULTADOS DE LA FUSION

Vías de corrección:

- Imputación no es posible.
- Ponderación con base a los datos del Multimedia Personal

EGM: 3ª OLA 2008 (Individuos)

	FUSIÓN	EGM Multimedia	EGM Radio	EGM Prensa	EGM Revistas	EGM Televisión
Población (000)	38.261	38.261	38.261	38.261	38.261	38.261
NACIONALIDAD						
España	35.491	33.145	35.293	34.807	33.929	34.165
Extranjero	2.770	5.116	2.968	3.454	4.332	4.096

Estudios telefónicos introducen un sesgo que va más allá del inmediato



RESULTADOS DE LA FUSION

Vías de corrección:

- Mayor peso de las edades extremas en el cálculo de las distancias.

Suavización de perfiles: Edades

EGM: 3ª OLA 2008 (Individuos)

Edad Promedio	EGM	
	FUSIÓN	Revistas
Población (000)	45,6	45,5
Ultimo periodo Revistas Semanales	43,7	43,4
El Jueves	36,7	35,7
Tiempo	46,6	47,0
Interviu	39,5	38,8
Actualidad Económica	38,8	40,0
Mi Cartera de Inversión	38,1	38,7
Motor 16	39,3	38,8
Autopista	37,6	36,1
Coche Actual	35,2	34,8
Auto Bild	33,4	33,0
Auto Hebdo Sport	33,8	33,0
Motociclismo	33,0	32,5
Solo Moto Actual	32,5	32,0
Nuevo Vale	29,3	25,8
Tp	43,3	43,0
Supertele	44,9	45,6
Telenovela	32,7	30,2
Hola	46,3	46,5
Semana	47,4	47,6
Lecturas	48,4	48,5
Diez Minutos	46,9	47,1
Pronto	48,2	48,3
Que Me Dices	42,6	42,4
Mia	45,7	45,6
Cuore	31,2	28,7

RESULTADOS DE LA FUSION

Suavización de perfiles: Edades

EGM: 3ª OLA 2008 (Individuos)

Edad Promedio	EGM	
	FUSIÓN	Revistas
Población (000)	45,6	45,5
Ultimo periodo Revista Mensuale	39,3	38,5
Digital +	39,6	38,7
Revista R	42,4	42,0
Ono	40,6	40,0
Cinemania	30,9	29,8
Fotogramas	37,2	36,9
Emprendedores	36,3	36,7
Capital	38,5	38,8
Que Leer	42,0	40,2
Speak Up	36,4	34,0
Play Boy	31,2	28,8
Man	33,4	31,9
Primera Línea	35,5	33,8
FHM	29,0	27,0
Sport Life	31,4	30,4
Men's Health	30,7	29,2
Runners World	35,1	33,4
Muy Interesante	38,9	38,1
Quo	37,5	37,0
Descubrir el Arte	38,9	38,5
La Aventura de la Historia	44,1	44,7
Clio	40,8	39,9
Historia y Vida	43,8	43,5
Historia National Geographic	37,9	37,0
Geo	41,5	41,0
Viajar	44,6	43,6
Rutas del Mundo	41,4	39,1
Deviajes	44,4	43,4
Viajes Nationalgeographic	35,9	34,4
Muchoviaje	39,0	38,0
National Geographic	39,8	39,0

	EGM	
	FUSIÓN	Revistas
	45,6	45,5
	39,3	38,5
Rolling Stone	28,6	26,5
La Revista 40	27,6	24,3
Investigación y Ciencia	42,4	42,5
Año Cero	40,0	40,5
Integral	48,1	47,3
Cuerpo-mente	44,2	43,9
Psicología Práctica	42,0	41,5
Avión Revue	37,5	38,1
Solo Camión	38,3	36,7
Transporte Mundial	39,0	36,4
Automóvil	39,3	39,1
Solo Auto 4X4	37,9	37,4
Auto Verde	36,6	36,9
Top Auto	34,8	33,8
Car And Driver	34,2	33,7
Autovía	33,8	32,9
Autofácil	37,2	36,1
Maxi Tuning	26,3	22,9
Racc Club	44,0	43,6
Marca Motor	34,0	32,9
Auto Bild 4x4	36,4	34,4
Todo Terreno	38,7	38,9
Solo Moto 30	32,2	30,3
La Moto	34,8	33,7
Moto Verde	31,9	29,2
Ciclismo a Fondo	38,6	37,7
Solo Bici	33,5	32,3
Bike	34,1	32,6
Micromanía	29,2	27,5
Hobby Consolas	26,5	23,6

	EGM	
	FUSIÓN	Revistas
	45,6	45,5
	39,3	38,5
Playmania	26,9	23,2
Computer Hoy Juegos	29,6	26,7
Playstation	29,3	25,5
Nintendo Acción	31,8	27,6
Personal Computer & Internet	37,6	37,5
Windows Vista	34,5	32,8
Pc Actual	36,9	37,2
Audio Vídeo Foto Hoy	35,8	37,0
Ragazza	31,7	28,3
Loka Magazine	24,2	19,5
Cosmopolitan	34,1	32,8
Elle	37,4	36,5
Vogue	35,1	34,3
Marie Claire	37,8	38,2
Telva	43,7	44,8
Woman	37,5	35,4
Clara	43,1	43,0
AR	44,8	44,3
Glamour	31,2	29,5
Instyle	33,3	32,9
Saber Vivir	46,9	46,9
Prevenir	53,0	55,8
Mente Sana	39,4	38,1
Psychologies	39,8	38,9
Aqua	36,9	33,7
Comer Bien Cada Día	42,7	42,3
Cocina Fácil	42,7	42,1
Cocina Ligera y Vida Sana	44,8	44,0
Patrones	43,9	42,3
Labores del Hogar	45,7	45,7

	EGM	
	FUSIÓN	Revistas
	45,6	45,5
	39,3	38,5
Mi Bebé y Yo	33,5	32,5
Ser Padres Hoy	37,2	36,3
Crecer Feliz	36,3	34,7
Tu Bebé	34,5	32,9
Guía del Niño	36,6	34,4
El Bebé de Ser Padres	34,9	33,6
Casa y Jardín	44,1	44,7
El Mueble	41,3	41,1
Nuevo Estilo	39,6	39,7
Micasa	40,4	40,3
Cosas de Casa	39,7	39,3
Casa Diez	40,6	39,8
Casa al Día	40,2	39,7
Casa & Campo	37,7	37,5
Mi Jardín Plantas en Casa	44,8	44,2
Interiores	42,7	42,7
Habitania	41,7	41,2
El Mueble Cocinas y Baños	37,9	37,8
Arquitectura y Diseño	39,8	39,8
El Mueble Casas de Campo	40,0	40,0
Casa Viva	38,0	37,6
Casa Joven	37,7	36,4
Ad	38,5	38,0
Bricolaje y Decoración	40,4	41,1
Trofeo Caza	37,5	36,4
Jara y Sedal	39,5	38,6
Trofeo Pesca	35,2	34,9
Jara y Sedal Pesca	37,5	35,9
Revista Oficial de la NBA	29,8	28,4

Niños en el hogar

RESULTADOS DE LA FUSION

EGM: 3ª OLA 2008 (Individuos)

	FUSIÓN	EGM Multimedia	EGM Radio	EGM Prensa	EGM Revistas	EGM Televisión
Población (000)	38.261	38.261	38.261	38.261	38.261	38.261
NIÑOS EN EL HOGAR						
CON NIÑOS	9.489	9.986	9.443	9.634	9.795	9.779
.De 0 a 12 meses	727	1.013	747	784	996	845
.De 13 a 24 meses	710	784	711	663	878	704
.De 25 a 36 meses	896	1.004	849	971	949	963
.De 3 a 4 años	1.807	1.852	1.780	1.836	1.880	1.840
.De 5 a 6	1.885	1.997	1.972	1.909	1.917	1.989
.De 7 a 8	1.885	1.824	1.796	1.897	1.878	1.874
.De 9 a 10	2.138	2.158	2.128	2.125	2.084	2.155
.De 11 a 13	3.135	3.293	3.152	3.166	3.173	3.233
SIN NIÑOS	28.772	28.275	28.818	28.627	28.466	28.482

Vías de corrección:

- Introducción presencia de niños en el cálculo de las distancias.



Duplicación medios

EGM: 3ª OLA 2008 (Individuos)

	FUSION (000)	MULTIMEDIA (000)	FUSION %	MULTIMEDIA %
Total	38261	38261	100,0%	100,0%
ESTRUCTURA DE MEDIOS U.P.				
Solo diarios	274	192	0,7%	0,5%
Solo Rev./Domin.	707	547	1,8%	1,4%
Solo radio	541	622	1,4%	1,6%
Solo Tv	4956	6897	13,0%	18,0%
Diaros y Revistas	515	374	1,3%	1,0%
Diaros y Radio	330	220	0,9%	0,6%
Diaros y TV	1898	1614	5,0%	4,2%
Revistas y Radio	881	584	2,3%	1,5%
Revistas y TV	4791	3929	12,5%	10,3%
Radio y TV	3939	4543	10,3%	11,9%
Diaros Revistas y Radio	823	604	2,2%	1,6%
Diaros Revistas y TV	3541	3034	9,3%	7,9%
Diaros Radio y TV	2587	2381	6,8%	6,2%
Revistas Radio y TV	5819	4945	15,2%	12,9%
Diaros Revistas Radio y TV	5946	6710	15,5%	17,5%
Ninguno	715	1067	1,9%	2,8%

El distinto valor de las audiencias brutas hace imposible que las duplicaciones sean las mismas que en el multimedia.



RESULTADOS DE LA FUSION

Audiencia de Prensa por Día de la Semana

EGM: 3ª OLA 2008 (hogares)

	Total			Sabado			Domingo			Sabado y Domingo			Lunes a Viernes		
	FUSION	FUSION LaV/S/D	EGM PRENSA	FUSION	FUSION LaV/S/D	EGM PRENSA	FUSION	FUSION LaV/S/D	EGM PRENSA	FUSION	FUSION LaV/S/D	EGM PRENSA	FUSION	FUSION LaV/S/D	EGM PRENSA
	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261	38.261
TOTAL DIARIOS	15.912	15.904	15.912	13.379	12.916	12.947	14.979	15.534	15.411	14.179	14.225	14.179	16.606	16.576	16.606
..De Pago T.	13.655	13.653	13.655	12.887	12.421	12.447	14.496	14.777	14.936	13.691	13.599	13.691	13.640	13.674	13.640
..Gratuitos T.	3.845	3.838	3.845	785	778	783	763	1.222	765	774	1.000	774	5.073	4.973	5.073
INFORMACIÓN GRAL.	14.291	14.274	14.291	11.410	10.945	11.010	13.210	13.711	13.609	12.310	12.328	12.310	15.083	15.053	15.083
..De Pago I.G.	11.809	11.798	11.809	10.873	10.410	10.467	12.692	12.906	13.098	11.782	11.658	11.782	11.819	11.854	11.819
..Gratuitos I.G.	3.845	3.838	3.845	785	778	783	763	1.222	765	774	1.000	774	5.073	4.973	5.073
DEPORTIVOS	4.250	4.238	4.250	4.196	4.049	4.035	3.974	4.129	4.136	4.085	4.089	4.085	4.315	4.297	4.315
ECONÓMICOS	271	270	271	314	286	295	105	112	124	209	199	209	296	298	296

Solución ensayada : Ponderación y Estratificación en base a
 PROVINCIA (50) X HABITAT (2) X DIA DE LA SEMANA (3)
 L-V / S / D

Software de planificación e investigación

Cambios en tamaño de archivos

	3er Acumulado 2007	3er Acumulado 2008
Número de registros	43.841	420.676
Tamaño ficheros MB (TOM Micro)	66,5	700
Tamaño ficheros MB (Galileo)	677	3860

Cambios en el rendimiento

	TOM Micro	
	3er Acumulado 2007	3er Acumulado 2008
Calculo plan de medios	5	42
Cálculo tabulación	3	22

	Galileo	
	3er Acumulado 2007	3er Acumulado 2008
Calculo plan de medios	97	880
Cálculo tabulación	2	23

El tiempo empleado en el cálculo aumenta aproximadamente de 7 a 10 veces respecto al empleado antes de la fusión



Nueva versión Galileo

- Disminución del tiempo en el cálculo de tabulaciones de aproximadamente un 25 %
- Creación del caché de variables en memoria para aumentar velocidad al recalcular informes.
- Eliminación de limitación en tamaño del fichero final (de 4GB a ilimitado)

Lanzamiento a mediados de 2009

CONCLUSIONES: LO APRENDIDO

Sistema que funciona.

Importancia del Multimedia:

- Única Fuente de muchos datos.
- Fuente de más “calidad”.
- Más Pequeña y más si se trata de amas.
- Más problemas para el equilibraje.

Muestra Multimedia y estratificación:

1968-1988 10000 / 200 = 50 entrevistas por estrato.

2008 10000 Mínimo.

Posibilidad de cuotas en las entrevistas telefónicas

Variables comunes:

- ¿sicográficos / tipologías?
- ¿consumo de medios?



CONCLUSIONES: LO APRENDIDO

Posibles Líneas de desarrollo:

- Línea actual bastante cerrada
 - Ponderación
 - Distancias
 - Estratificación
- Estratificación ad hoc muestral vs poblacional

Tratamiento de fuentes asimétricas (AIMC Marcas):

- Muestreo
 - Periodicidad
- Solo donantes / no receptores
Información multinivel
Estratos solapados

- Modelos probabilísticos bayesianos ...

