

Departamento Administrativo Nacional de Estadística

Dirección de Censos y Demografía



**METODOLOGIA
ESTIMACIONES Y PROYECCIONES DE POBLACIÓN , POR ÁREA,
SEXO Y EDAD PARA
LOS DOMINIOS DE
LA GRAN ENCUESTA INTEGRADA
DE HOGARES
(2000 – 2009)**

Julio de 2008

Contenido

<u>INTRODUCCION</u>	3
<u>I. PROYECCIONES DE POBLACIÓN POR ÁREA, EDAD Y SEXO, A NIVEL NACIONAL 2005 - 2020</u>	4
1.1 MODELO DE AREAS MAYORES	5
1.1.1 PROYECCIÓN DE LA POBLACIÓN POR SEXO Y GRUPOS QUINQUENALES DEL RESTO NACIONAL	6
1.1.2 ESTIMACIÓN DE LA FECUNDIDAD PARA EL RESTO NACIONAL	6
1.1.3 ESTIMACIÓN DEL SALDO NETO MIGRATORIO TOTAL DEL RESTO NACIONAL	7
1.1.4 ESTIMACIÓN DE LA MORTALIDAD PARA EL RESTO NACIONAL	7
<u>II. PROYECCIONES DE POBLACIÓN POR ÁREA, EDAD Y SEXO, A NIVEL DEPARTAMENTAL 2005 – 2020.</u>	8
<u>III. PROYECCIONES DE POBLACIÓN POR ÁREA, EDAD Y SEXO, A NIVEL MUNICIPAL 2005 – 2009.</u>	12
<u>IV- ESTIMACION DE LA POBLACIÓN POR ÁREA, EDAD Y SEXO, PARA LOS DOMINIOS DE LA GEIH, 2000 – 2005.</u>	21

INTRODUCCION

Los resultados del Censo 2005 evidencian una dinámica demográfica diferente a la considerada en las últimas proyecciones, en cada uno de los entes territoriales, lo cual se refleja en la distribución espacial de la población, su estructura etárea, distribución por sexo y en los principales indicadores determinantes de la dinámica poblacional. Por tener dichos cambios, efectos sobre la distribución espacial de la población y sus estructuras, es básico considerarlos en los procesos de expansión que se llevan a cabo en la Gran Encuesta Integrada de Hogares.

Bajo este contexto, las proyecciones de población deben garantizar la coherencia de las estimaciones entre los niveles nacional, departamental y municipal considerando el uso de esta información en la decisión de política pública sobre distribución de recursos de inversión sectorial, evaluación de política en salud y educación, así como en la parte laboral.

Por todo lo anterior, el DANE ha implementado una serie de adecuaciones metodológicas para la elaboración de las proyecciones, con el fin de garantizar la mayor robustez técnica en los modelos de acuerdo con los niveles territoriales, dado que a medida que el nivel de desagregación es menor, los insumos de información disponibles, presentan deficiencias en su calidad.

En este orden de ideas, en éste documento se detalla la nueva metodología utilizada para obtener los insumos de población por quincenas, para el periodo 2000 – 2009 desagregados por sexo, edad a nivel total, cabecera y resto de cada uno de los dominios de la encuesta. Este modelo metodológico garantiza la coherencia de las estructuras poblacionales de dichos dominios, con los niveles departamental y nacional ya publicados por el DANE.

I. PROYECCIONES DE POBLACIÓN POR ÁREA, EDAD Y SEXO, A NIVEL NACIONAL 2005 - 2020

Como resultado de la evaluación del Método de Cohortes aplicado para obtener las proyecciones de las cabeceras departamentales y por diferencia las del resto, se encontró que éstas últimas en varios departamentos y Bogotá mostraban unas distribuciones no compatibles desde el punto de vista de tendencias demográficas para los periodos finales de la proyección. Por este motivo, se planteó un nuevo proceso que consiste en proyectar el área resto por Cohortes y obtener la cabecera por diferencia.

Para realizar lo anterior, se requiere obtener como insumo para el nivel nacional las tasas globales y específicas de fecundidad, así como la relación de sobrevivencia al nacer. Considerando la falta de información para construir tablas de vida para ésta área, se utilizó un modelo demográfico denominado de Áreas Mayores que permite estimar por diferencias de dos áreas jerárquicas mayores proyectadas por componentes, en éste caso total y cabecera, las componentes demográficas del área resto y por lo tanto la estructura por edad y sexo calibrada para proyectar por Cohortes el resto de cada departamento.

Como se indica, el total nacional se proyectó por el Método de las Componentes¹, resultados que se utilizan en la aplicación de la metodología propuesta. Para el caso de la cabecera a nivel nacional, también se utilizó la misma metodología, lo cual permite obtener los resultados de los indicadores demográficos que se presentan en la tabla 1.

Tabla No. 1
Colombia – Cabecera: Principales Indicadores Demográficos 2005 - 2020

PERIODO	TASAS MEDIAS ANUALES DE CRECIMIENTO (%)		TASAS IMPLICITAS (POR MIL)			NUMERO ESTIMADO DE		MIGRANTES NETOS	
	EXPONENCIAL	GEOMETRICO	CRECIMIENTO NATURAL	BRUTA NATALIDAD	BRUTA MORTALIDAD	NACIMIENTOS	DEFUNCIONES	TOTAL	TASA (POR MIL)
2005-2010	1.51	1.52	12.99	18.43	5.44	3,053,764	901,136	335,817	2.03
2010-2015	1.41	1.42	12.24	17.36	5.13	3,094,191	913,856	315,671	1.77
2015-2020	1.27	1.28	11.15	16.40	5.24	3,123,414	998,629	284,103	1.49

¹ Para los aspectos metodológicos ver: DANE, Proyecciones nacionales y departamentales de población 2006 – 2020. Septiembre de 2007.

AÑO	TASAS DE REPRODUCCION		TASAS DE FECUNDIDAD (POR 1,000 MUJERES)		EDAD MEDIA DE LA FECUNDIDAD
	BRUTO	NETO	GLOBAL	GENERAL	(AÑOS)
2005-2010	1.05	1.02	2150.5	65.9	26.48
2010-2015	1.01	0.99	2074.5	63.3	26.44
2015-2020	0.99	0.98	2039.5	61.7	26.42

AÑOS	ESPERANZA DE VIDA AL NACER (AÑOS)			TASA MORTALIDAD	DEFUNCIONES		
	HOMBRES	MUJERES	TOTAL	INFANTIL	EDAD 0	EDAD 0-4	EDAD 1-4
2005-2010	71.73	78.50	75.02	15.1	46,252	49,879	3,626
2010-2015	74.26	80.48	77.29	14.0	43,331	44,735	1,404
2015-2020	75.50	81.72	78.53	12.3	38,540	39,024	484

Fuente: Censo 2005. Estimación a través del Paquete PRODEM

Teniendo en cuenta los resultados nacionales, así como los presentados en la tabla anterior, se aplicó el Modelo de Áreas Mayores para obtener los indicadores demográficos y la estructura etárea del área resto. Los aspectos metodológicos de éste modelo se presenta a continuación.

1.1 MODELO DE AREAS MAYORES

Es un método demográfico que requiere disponer de las proyecciones de dos áreas jerárquicas mayores obtenidas por el método de las componentes, para lo cual se utiliza el paquete PRODEM (Naciones Unidas:1982).

El objetivo de este proceso es obtener por diferencia el Resto Nacional, para lo cual se requiere tener los siguientes insumos:

1. Las proyecciones por sexo y edad del periodo 2005 – 2020 por componentes total nacional, cuyos resultados corresponden a los obtenidos en la primera fase del proceso (DANE: 2007). Para efectos de este procedimiento, se denomina Área Geográfica del Nivel Jerárquico Superior la cual se denota con la sigla JS. Se debe ingresar los resultados de PRODEM en un archivo independiente, los cuales durante el proceso la aplicación toma los insumos necesarios.
2. Los insumos para la cabecera nacional (2005), se organizaron por edad y sexo, así como los niveles y estructuras de las componentes de fecundidad, mortalidad y migración para aplicar el método de las componentes y

obtener los insumos requeridos por el modelo del Área Mayor que denotaremos como AM. Para éste proceso se aplicaron los mismos instrumentos metodológicos y computacionales utilizados en la proyección del total nacional.

Los componentes del crecimiento demográfico para el Resto Nacional, se obtienen por diferencia de los insumos anteriores, utilizando los algoritmos que a continuación se explican:

1.1.1 Proyección de la población por sexo y grupos quinquenales del Resto Nacional

Se obtiene por diferencia entre la población del nivel Jerárquico Superior (NJS) y la del Área Mayor (NAM) para el mismo grupo quinquenal de edades y sexo en el instante t, es decir:

$$NR_{x,x+5}^{Sexo,t} = NJS_{x,x+5}^{Sexo,t} - NAM_{x,x+5}^{Sexo,t}$$

1.1.2 Estimación de la fecundidad para el Resto Nacional

Para la obtención de las TGF y las Específicas para cada quinquenio de la proyección, se deben estimar en primer lugar los nacimientos por edad de la madre del resto (BR) para cada grupo de edad entre 15 y 49 años, mediante la siguiente expresión:

$$BR_{x,x+5}^{t,t+5} = \frac{5[fJS_{x,x+5}^{t,t+5} * (NFJS_{x,x+5}^t + NFJS_{x,x+5}^{t+5}) - fAM_{x,x+5}^{t,t+5} * (NFAM_{x,x+5}^t + NFAM_{x,x+5}^{t+5})]}{2}$$

Entonces las tasas específicas de fecundidad se obtiene como:

$$fAR_{x,x+5}^{t,t+5} = \frac{[\frac{1}{5} * BR_{x,x+5}^{t,t+5}]}{[\frac{1}{2} * (NFR_{x,x+5}^t + NFR_{x,x+5}^{t+5})]}$$

$$TGF^{t,t+5} = 5 \sum_{\substack{t=2000 \\ x=15}}^{\substack{t=2015 \\ x=45}} fAR_{x,x+5}^{t,t+5}$$

Con . $fJS_{x,x+5}^{t,t+5}$ = Tasas de fecundidad del Total Nacional

$fAR_{x,x+5}^{t,t+5}$ = Tasas de fecundidad del Resto Nacional

$NFJS_{x,x+5}^t$ = Población femenina media en edad fértil del Total Nacional

$NFAM_{x,x+5}^t$ = población femenina media en edad fértil de la Cabecera Nacional

Evaluada la estructura de la fecundidad resultante, se encontró que su tendencia es completamente incoherente. Por lo tanto, se realizó un ajuste teniendo en cuenta las estructuras modelo de fecundidad de Naciones Unidas (Naciones Unidas: 1965). El resultado del proceso determinó que para el Resto de Colombia, la estructura que mejor se adapta es la de alta fecundidad, tipo A, teniendo en cuenta la coherencia con la estructura total encontrada en las proyecciones por componentes de éste nivel.

1.1.3 Estimación del Saldo Neto Migratorio Total del Resto Nacional

Para la proyección de cabecera nacional, se realizó la estimación del componente de migración interna con el fin de obtener el saldo neto migratorio total, es decir internacional más interno. Esto permite que la migración de resto contenga el efecto de la emigración interna hacia cabecera, con lo cual se puede obtener el saldo neto migratorio de la siguiente forma:

$$SNMR_{x,x+5}^{t+5} = SNMJS_{x,x+5}^{t+5} - SNMAM_{x,x+5}^{t+5}$$

1.1.4 Estimación de la mortalidad para el Resto Nacional

El modelo estima en primer lugar la relaciones de sobrevivencia y luego calcula en función de éstas otros indicadores como la esperanza de vida al nacer por sexo y total. La relación de sobrevivencia por sexo y edad, a partir de los 5 años de edad se obtiene mediante la expresión:

$$PR_{x,x+5}^{t,t+5} = \frac{NR_{x+5,5}^{t,t+5} - SNMR_{x+5,5}^{t,t+5}}{NR_{x+5,5}^t}$$

Para el caso del grupo 0 a 4 años, se estima la sobrevivencia al nacer que aplicada a los nacimientos permite estimar la población de dicho grupo. Esta relación se obtiene de la siguiente forma:

$${}_5PR_b^{t,t+5} = \frac{NR_{0,6}^{t+5} - SNMR_{0,4}^{t+5}}{BR^{t,t+5}}$$

II. PROYECCIONES DE POBLACIÓN POR ÁREA, EDAD Y SEXO, A NIVEL DEPARTAMENTAL 2005 – 2020.

Teniendo en cuenta la inestabilidad del área resto en la mayoría de los departamentos, cuando se obtienen por diferencia entre el total conciliado y la cabecera obtenida por el Método de Cohortes, se cambió el proceso de estimación. Este cambio consiste en obtener primero por el Método de Cohortes² la parte resto de cada departamento a partir de los insumos nacionales como techo, con lo cual se tiene en cuenta de acuerdo con las estructuras calibradas de los censos 1993 y 2005, cuya combinación permite considerar la dinámica diferencial existente en cada resto departamental; la cabecera se obtiene por diferencia garantizando la calibración al total municipal y generando un proceso final de calibración para garantizar que la sumatoria de municipios por departamento sea igual al total departamental definido.

Como resultado del análisis de los resultados al aplicar el Método de Cohortes en el área resto departamental, fue necesario ajustar los factores K de crecimiento de cada cohorte y sexo, con lo cual se obtuvo la tendencia y nivel más coherente para cada una de las cohortes por sexo y, por lo tanto los indicadores demográficos adecuados para el periodo de proyección que para este niveles corresponde al intervalo 2005 – 2020.

Para elaborar las proyecciones cabecera-resto departamentales se utilizó el aplicativo PeqAr desarrollado por el Instituto Brasileiro de geografia y Estadística IBGE, el cual recoge la metodología de relación de cohortes y adicionalmente, permite la obtención de edades simples por sexo para cada área proyectada. Los insumos requeridos en este proceso son: Relaciones de sobrevivencia al nacer, tasas específicas de fecundidad Proyecciones de cabecera resto nacionales y las estructuras censales ajustadas por área 1993 y 2005.

Es importante señalar que para la obtención del área geográfica “resto”, por ser más pequeña que las de la cabecera y más volátil se realizó por el Método de Cohortes con el fin de garantizar la coherencia de las estructuras de cada departamento con el nacional. La obtención de cabecera se hizo restándole a la

² DANE. Proyecciones Municipales Cabecera-Resto por sexo y edad 2006-2020. Capítulo 3. 2008

estructura total el valor proyectado de resto; asimismo, lo que se obtuvo con PeqAr fue principalmente las estructuras en términos porcentuales ya que los totales de población estaban calculados con anterioridad. Las diferencias resultantes entre la suma de los departamentos y el nacional fueron distribuidas proporcionalmente al tamaño de cada departamento.

Entre los instrumentos utilizados en esta etapa se tienen la plantilla desarrollada en Excel llamadas **00 Nacional resto área menor.xls**, la cual permite analizar y dar parámetros para los ajustes de las salidas del programa PeqAr (factor K), asimismo, admite la imputación de estas estructuras coherentes a los niveles de población preestablecidos. Entre los criterios para evaluar la consistencia de estos resultados se tienen pirámides de población, índices de masculinidad, % de grandes grupos de edad, relaciones de dependencia, relación niños mujeres, entre otros.

El proceso de imputación de estructuras a los niveles ya definidos se aplica de la siguiente manera: Se calculan los coeficientes de ajuste ${}_5Coef_x^{2005}$ para cada grupo de edad dividiendo el valor de cada grupo de edad ${}_5GE_x^{2005}$ por la suma de todos los grupos de tal manera que la suma de estos coeficientes siempre es igual a 1, garantizando que dichos niveles ya definidos siempre se respetan. Una vez calculados esos coeficientes, se multiplican por el nivel ya establecido quedando así la misma estructura de población obtenida con PeqAr, pero ajustado al nivel previamente calculado y publicado.

De forma general se puede expresar que el coeficiente de ajuste se calcula así:

$${}_5Coef_x^{2005} = \frac{({}_5GE_x^{2005})}{\sum_{X=0-4}^{80ymas} ({}_5GE_x^{2005})} \text{ donde:}$$

${}_5Coef_x^{2005}$ es el coeficiente de ajuste para la edad x

$({}_5GE_x^{2005})$ es el grupo de edad x resultante del modelo PeqAr (relación de cohortes)

$\sum_{X=0-4}^{80ymas} ({}_5GE_x^{2005})$ es la suma de todos los grupos de edad obtenido de PeqAr

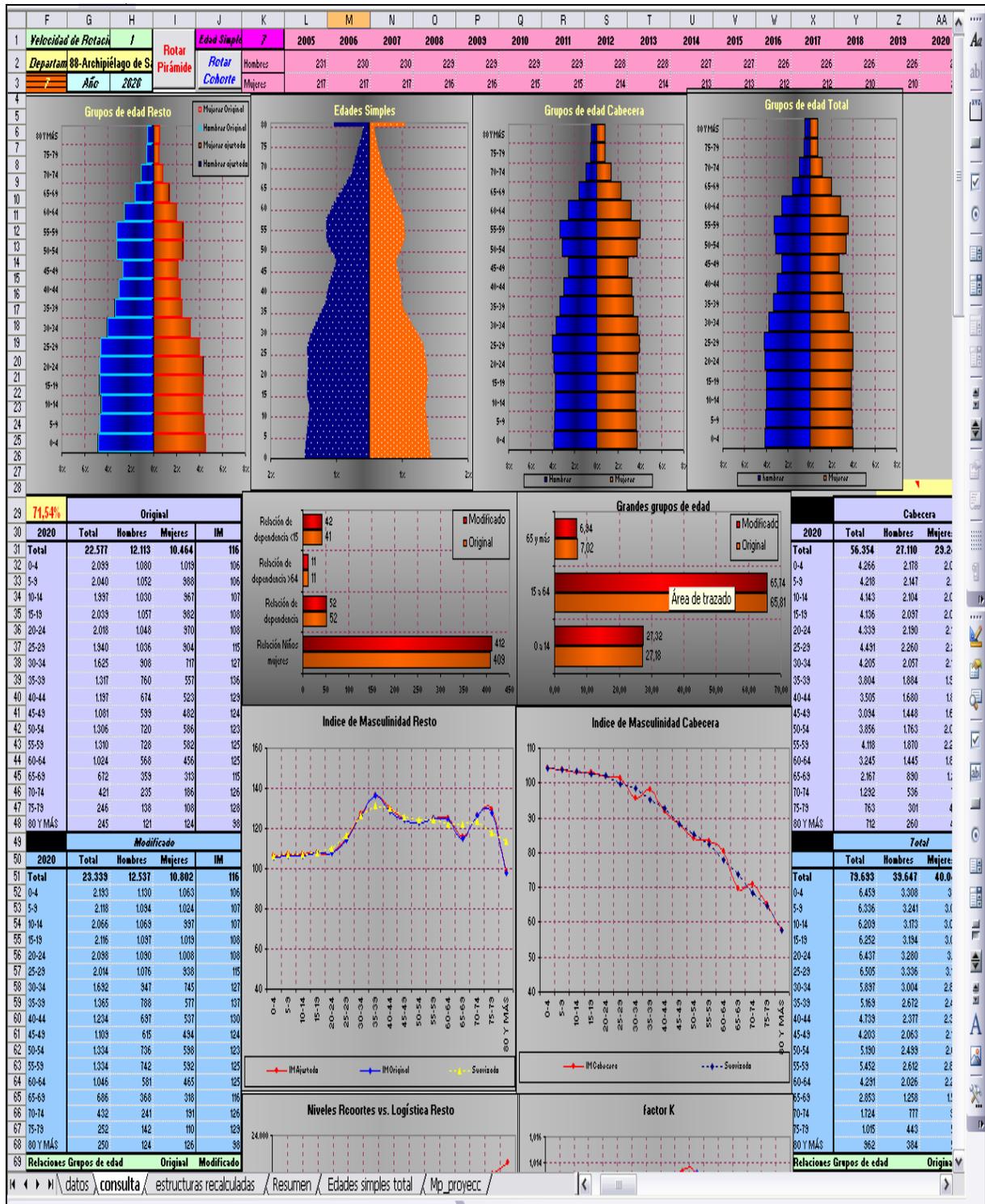
La estructura ajustada con niveles predefinidos se obtiene con el algoritmo

$${}_5GE_x^{2005} = {}_5Coef_x^{2005} * Niv_pred \text{ en donde}$$

${}_5GE_x^{2005}$ es el grupo de edad x ajustado, acorde con los niveles definidos con anterioridad

Niv_pred es el nivel previamente calculado o techo municipal

Figura 1. Vista de la hoja consulta de la plantilla 00 Nacional resto área menor.xls,



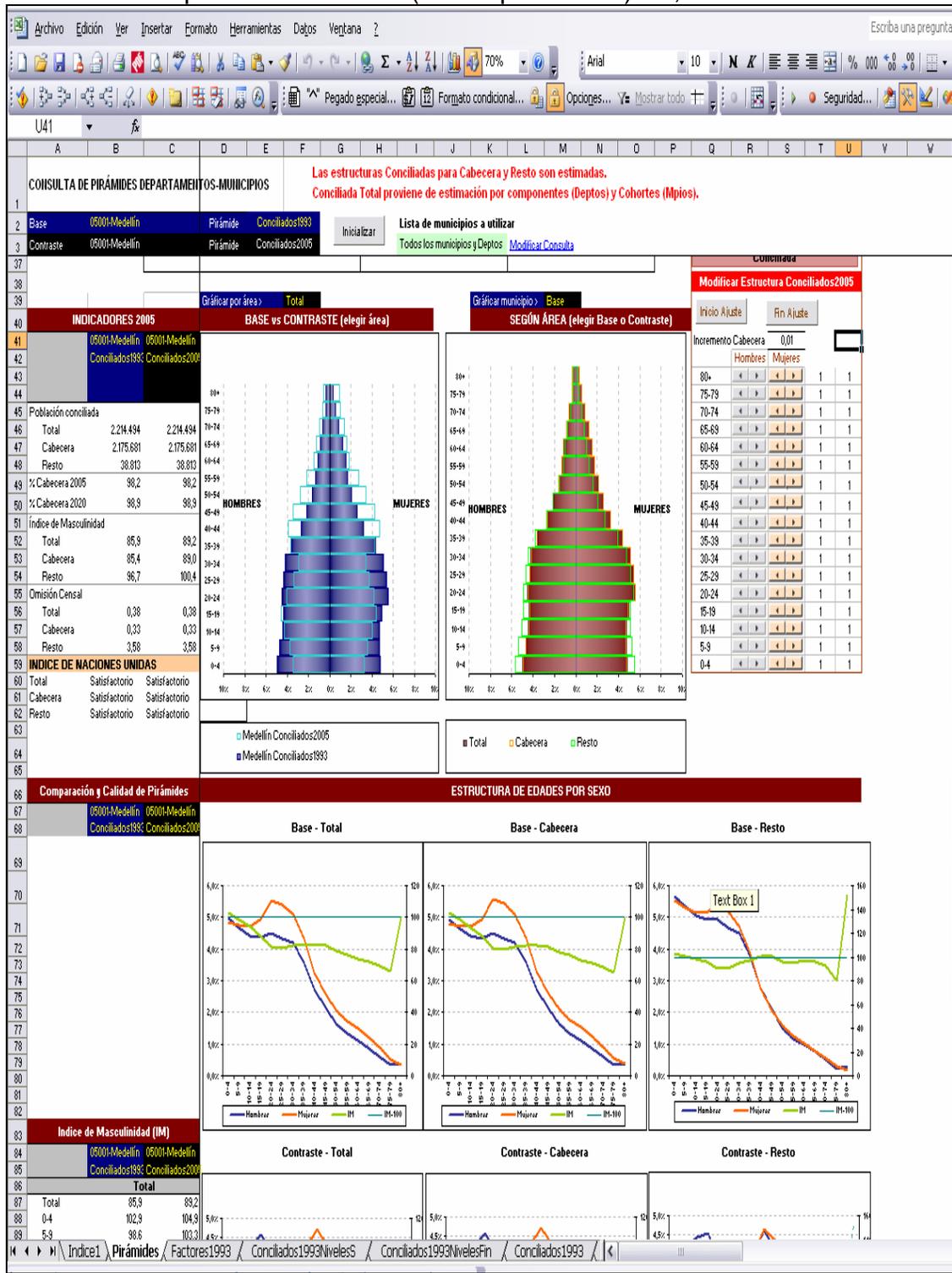
III. PROYECCIONES DE POBLACIÓN POR ÁREA, EDAD Y SEXO, A NIVEL MUNICIPAL 2005 – 2009.

Para la elaboración de las proyecciones municipales por área, no se cuenta con los insumos necesarios para utilizar el método de los componentes ni el de relación de cohortes, lo cual implicó desarrollar una estructura teórica cabecera resto 2020, teniendo como elementos las proyecciones departamentales cabecera resto para el 2020 y las proyecciones de población por grupos de edad para los municipios total 2020.

Esta etapa se puede dividir en tres sub etapas, a saber:

- a. **Estimación de estructuras cabecera - resto 2005.** Para este proceso se tuvo en cuenta las estructuras censales por área, se hizo un proceso de suavización de índices de masculinidad de la cabecera, el resto se calculó por diferencia entre el total municipal y la cabecera. En esta sub etapa se utilizó la plantilla **Estimación Cabecera Conciliada Municipios v3 Excel2003** (con deptos modif).xls, en donde es posible ajustar en las cabeceras cada grupo de edad y sexo de manera que se pueden obtener estructuras coherentes demográficamente de acuerdo con indicadores tales como índice de masculinidad, Índice de Naciones Unidas entre otros. Una vez halladas estas estructuras es necesario distribuir las diferencias resultantes de la suma municipal por edades y el departamento de manera proporcional al tamaño de cada municipio. En esta parte se recurrió a la plantilla **Tabla Calibre prorateo.xls**, la filosofía de esta es bastante simple: Al seleccionar el departamento a calibrar, aparecen todos los municipios que lo conforman con sus respectivas estructuras de población por grupos de edad y sexo; aparecen las diferencias (suma municipios menos conciliado departamental) a distribuir entre todos de manera proporcional a su peso dentro de su respectivo grupo, se minimizan dichas diferencias y se utiliza al final un comodín el cual va a absorberlas cuidando de no alterar su estructura original. El resultado son estructuras de población por grupos de edad total, cabecera y resto armonizadas con los valores departamentales.

Figura 2. Vista de la hoja pirámides de la plantilla Estimación Cabecera Conciliada Municipios v3 Excel2003 (con depts modif).xls,



El prorrateo tiene como fin eliminar las diferencias existentes al sumar todos los municipios con sus respectivos grupos de edad y compararlos con la correspondiente estructura departamental al cual pertenecen esos municipios. Este algoritmo se puede formular a través de la expresión

$${}_5\Delta_x = {}_5GE_depto_x^{2005} - \sum {}_5GE_x \text{ donde:}$$

${}_5\Delta_x$ es la diferencia del grupo de edad x departamental y la suma de ese mismo grupo de todos los municipios.

${}_5GE_depto_x^{2005}$ es el grupo de edad x del departamento

$\sum {}_5GE_x$ es la suma de todos los municipios del grupo de edad x .

Cuando ${}_5\Delta_x > 0$, significa que los municipios están subvalorados en el grupo de edad x , por lo tanto, esas diferencias se deben sumar de manera proporcional al tamaño de cada municipio, en el caso contrario (diferencias < 0), se deben restar ya que estos grupos de edad están sobrevalorados. La proporción para redistribuir esas diferencias es la misma participación o peso relativo del grupo de edad x del municipio n , sobre la suma de ese grupo de edad de todos los municipios que pertenecen al departamento, es decir:

$${}_5Peso_ \%_x = \frac{({}_5GE_x)_i}{\sum_{i=1}^n ({}_5GE_x)_i}$$

Donde ${}_5Peso_ \%_x$ es la participación del grupo de edad x del municipio i respecto a la suma de todos grupos de edad x de todos los municipios que componen un departamento

$({}_5GE_x)_i$ es el grupo de edad x del municipio n ,

y $\sum_{i=1}^n ({}_5GE_x)_i$ es la suma de todos los grupos de edad x de los municipio de i hasta n

Una vez hallada esa proporción, se multiplica por ${}_5\Delta_x$, cuyo resultado es sumado (${}_5\Delta_x > 0$) o restado (${}_5\Delta_x < 0$) al grupo de edad x del municipio i .

$({}_5GE_aj_x) = ({}_5GE_x)_i + ({}_5Peso_ \%_x * {}_5\Delta_x)$, teniendo en cuenta que

$({}_5GE_aj_x)$ es el grupo de edad x ajustado del municipio i .

Figura 3. Vista de la hoja consulta de la plantilla *Tabla calibre prorrateor.xls*,

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?																		
	G	H	I	J	K	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Año a ajustar	Otro	Urbanismo		54.9%													
2	Departamento	La Guajira	Actualizar															
3			Suma Municipios															
4	Iniciar	La Guajira	Total			Resto			Total			Cabecera			Resto			Total
5	Ajustes a estructura	Grupos de	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total
6	2006-2020	Total	1.067.126	528.307	538.819	481.259	250.079	231.180	1.067.126	528.307	538.819	585.867	278.269	307.598	481.259	250.038	231.221	0
7	tecnica	0-4	134.108	68.506	65.602	66.125	34.018	32.107	134.108	68.506	65.602	67.225	34.119	33.106	66.883	34.387	32.496	0
8	2020	5-9	126.043	64.302	61.741	58.892	30.254	28.628	126.043	64.302	61.741	67.255	34.099	33.156	58.788	30.203	28.595	0
9	Ajusta 2020	10-14	114.928	58.498	56.430	56.198	28.803	27.395	114.928	58.498	56.430	60.992	30.811	30.181	53.936	27.687	26.249	0
10		15-19	104.000	52.625	51.375	55.725	28.582	27.143	104.000	52.625	51.375	47.067	23.413	23.654	56.933	28.212	27.721	0
11		20-24	94.036	47.107	46.899	46.068	23.670	22.398	94.036	47.107	46.899	47.231	23.108	24.123	46.805	24.029	22.776	0
12		25-29	85.887	42.733	43.154	38.056	19.893	18.163	85.887	42.733	43.154	47.889	22.898	24.791	38.188	19.835	18.363	0
13		30-34	83.468	41.057	42.411	34.257	18.329	15.928	83.468	41.057	42.411	48.921	22.625	26.296	34.547	18.432	16.116	0
14		35-39	66.155	31.445	34.710	24.929	13.030	11.899	66.155	31.445	34.710	41.097	18.301	22.796	25.058	13.144	11.914	0
15		40-44	56.339	26.463	29.876	21.854	11.227	10.427	56.339	26.463	29.876	34.874	15.324	19.550	21.465	11.109	10.326	0
16		45-49	45.733	21.560	24.173	18.046	9.462	8.584	45.733	21.560	24.173	27.864	12.165	15.639	17.869	9.395	8.474	0
17		50-54	40.569	19.036	21.533	16.168	8.533	7.635	40.569	19.036	21.533	24.562	10.581	13.981	16.007	8.495	7.552	0
18		55-59	33.442	15.995	17.447	13.040	7.061	5.979	33.442	15.995	17.447	20.507	8.974	11.533	12.935	7.021	5.914	0
19		60-64	26.815	12.823	13.992	10.571	5.753	4.888	26.815	12.823	13.992	16.343	7.097	9.246	10.472	5.726	4.746	0
20		65-69	18.872	9.202	9.670	7.444	4.132	3.312	18.872	9.202	9.670	11.506	5.115	6.391	7.366	4.087	3.279	0
21		70-74	15.178	7.247	7.931	5.996	3.263	2.723	15.178	7.247	7.931	9.161	3.985	5.186	5.997	3.262	2.735	0
22		75-79	10.193	4.758	5.435	3.825	2.055	1.770	10.193	4.758	5.435	6.476	2.762	3.714	3.717	1.996	1.721	0
23		80 Y MÁS	11.380	4.920	6.440	4.275	2.004	2.270	11.380	4.920	6.440	7.077	2.892	4.185	4.283	2.028	2.255	0
24			1	2	3	7	8	9										
25			Total	Hombres	Mujeres	Total	Hombres	Mujeres	Factor diferencias	Originales mas diferencias	factor prorrateo	PRORRATED						
26	44001	Riohacha	295.619	145.761	149.858	43.844	25.523	18.321		43.878	25.543	18.335		43.844	25.523	18.321		
27	0-4		36.624	18.778	17.846	5.107	2.417	2.690	0,07028819	0,082779419	0,0549	2,391	2,658	0,0545	0,0606	5,046	2,390	2,656
28	5-9		35.668	18.307	17.361	4.717	2.487	2.230	0,08234204	0,078012944	4,725	2,492	2,233	0,0568	0,0508	4,721	2,490	2,231
29	10-14		33.421	17.162	16.259	5.478	2.883	2.595	0,10428291	0,098860909	5,707	2,989	2,708	0,0683	0,0617	5,703	2,987	2,706
30	15-19		28.704	14.511	14.193	6.024	3.343	2.681	0,11439272	0,096710683	5,896	3,271	2,626	0,0745	0,0598	5,891	3,268	2,623
31	20-24		23.866	11.823	12.043	3.677	2.215	1.462	0,092190282	0,064190376	3,620	2,182	1,438	0,0497	0,0328	3,617	2,180	1,437
32	25-29		22.691	11.018	11.673	2.837	1.649	1.188	0,083358671	0,064895311	2,829	1,654	1,175	0,0377	0,0288	2,827	1,653	1,174
33	30-34		23.953	11.538	12.415	3.311	2.070	1.241	0,112304688	0,077089988	3,285	2,058	1,227	0,0469	0,0380	3,282	2,056	1,226
34	35-39		18.737	9.009	10.728	2.353	1.510	843	0,114881315	0,070757082	2,339	1,497	842	0,0341	0,0192	2,337	1,496	841
35	40-44		16.231	7.372	8.859	2.034	1.324	710	0,108861657	0,068758474	2,051	1,334	717	0,0304	0,0183	2,049	1,333	716
36	45-49		12.778	5.939	6.839	1.754	1.164	590	0,123895689	0,069624734	1.770	1.172	598	0,0287	0,0136	1.769	1.171	598
37	50-54		11.047	5.059	5.989	1.799	1.108	691	0,131046718	0,091489941	1.817	1.118	699	0,0255	0,0159	1.815	1.117	698
38	55-59		9.324	4.530	4.794	1.465	996	469	0,141860134	0,07931	1.002	474	0,0228	0,0108	1.475	1.001	474	
39	60-64		7.227	3.501	3.726	1.217	894	323	0,156123934	0,0880	898	328	0,0205	0,0075	1.225	897	323	
40	65-69		5.138	2.657	2.481	882	656	226	0,160508931	0,06993	663	228	0,0151	0,0052	890	662	228	
41	70-74		3.894	1.981	1.913	706	524	182	0,160637646	0,06654479	705	524	181	0,0119	0,0041	705	524	181
42	75-79		2.639	1.341	1.298	330	199	131	0,099693939	0,076118536	340	205	135	0,0047	0,0031	340	205	135
43	80 Y MÁS		2.677	1.235	1.442	163	94	69	0,04142018	0,03059867	162	83	69	0,0019	0,0016	162	83	69
44	44035	Albania	28.474	15.035	13.439	14.714	8.085	6.629		14.706	8.082	6.624		14.714	8.090	6.624		
45	0-4		3.589	1.838	1.750	1.921	992	929	0,02849111	0,028588134	1.899	981	918	0,0667	0,0624	1.906	988	918
46	5-9		3.276	1.700	1.576	1.656	866	790	0,028672648	0,027636872	1.659	868	791	0,0590	0,0538	1.659	868	791
47	10-14		2.943	1.544	1.405	1.617	846	771	0,030595857	0,029372548	1.685	880	805	0,0598	0,0547	1.685	880	805
48	15-19		2.812	1.472	1.340	1.938	1.001	937	0,03426674	0,033801089	1.896	979	917	0,0666	0,0624	1.897	980	917
49	20-24		2.720	1.420	1.300	1.415	739	676	0,030754505	0,029680365	1.393	728	665	0,0495	0,0452	1.393	728	665
50	25-29		2.420	1.270	1.150	1.174	629	545	0,031711621	0,029679246	1.170	631	539	0,0429	0,0387	1.170	631	539
51	30-34		2.306	1.223	1.083	1.080	639	441	0,034867969	0,027385808	1.071	635	436	0,0432	0,0396	1.071	635	436
52	35-39		2.019	1.081	938	845	508	337	0,038648813	0,02828605	841	504	337	0,0343	0,0229	841	504	337
53	40-44		1.656	887	769	729	433	296	0,03887243	0,028665505	735	436	299	0,0296	0,0203	735	436	299
54	45-49		1.235	655	580	565	328	237	0,034912887	0,027987902	570	330	240	0,0224	0,0163	570	330	240
55	50-54		1.065	575	490	509	298	211	0,035245447	0,027939619	514	301	213	0,0205	0,0145	514	301	213
56	55-59		856	466	390	419	252	167	0,035882323	0,028238079	422	253	169	0,0172	0,0115	422	253	169
57	60-64		696	406	290	395	265	130	0,046280126	0,027914688	398	266	132	0,0181	0,0130	398	266	132
58	65-69		349	209	140	146	99	47	0,024223147	0,014333638	147	100	47	0,0089	0,0032	147	100	47
59	70-74		221	129	92	123	81	42	0,024831392	0,015356449	123	81	42	0,0055	0,0029	123	81	42
60	75-79		131	71	60	68	39	29	0,019539078	0,016950668	70	40	30	0,0027	0,0020	70	40	30
61	80 Y MÁS		174	88	86	114	70	44	0,034516795	0,019512195	113	69	44	0,0047	0,0030	113	69	44
62	44078	Barrancas	37.554	18.867	18.687	17.877	9.700	8.177										

- b. Estimación de la estructura teórica 2020.** Con las estructuras calibradas cabecera y resto 2005, la proyección de población por grupos de edad y sexo departamental 2020 y municipal total, por sexo y edad 2020, se procede a estimar una posible estructura de cabecera para este último año, teniendo como supuesto que las proporciones del resto de los municipios guardan la misma relación del departamento.

De manera formal, se puede expresar que la estructura teórica para el año 2020 se efectúa a partir del algoritmo:

$${}_5MN_x^{2020} = \frac{({}_5MN_x^{2005} * {}_5DP_x^{2020})}{{}_5DP_x^{2005}}$$

Donde:

${}_5MN_x^{2020}$ es el grupo x de edad teórico del año 2020 de la cabecera municipal

${}_5MN_x^{2005}$ Es el grupo de edad x de la cabecera municipal en el 2005

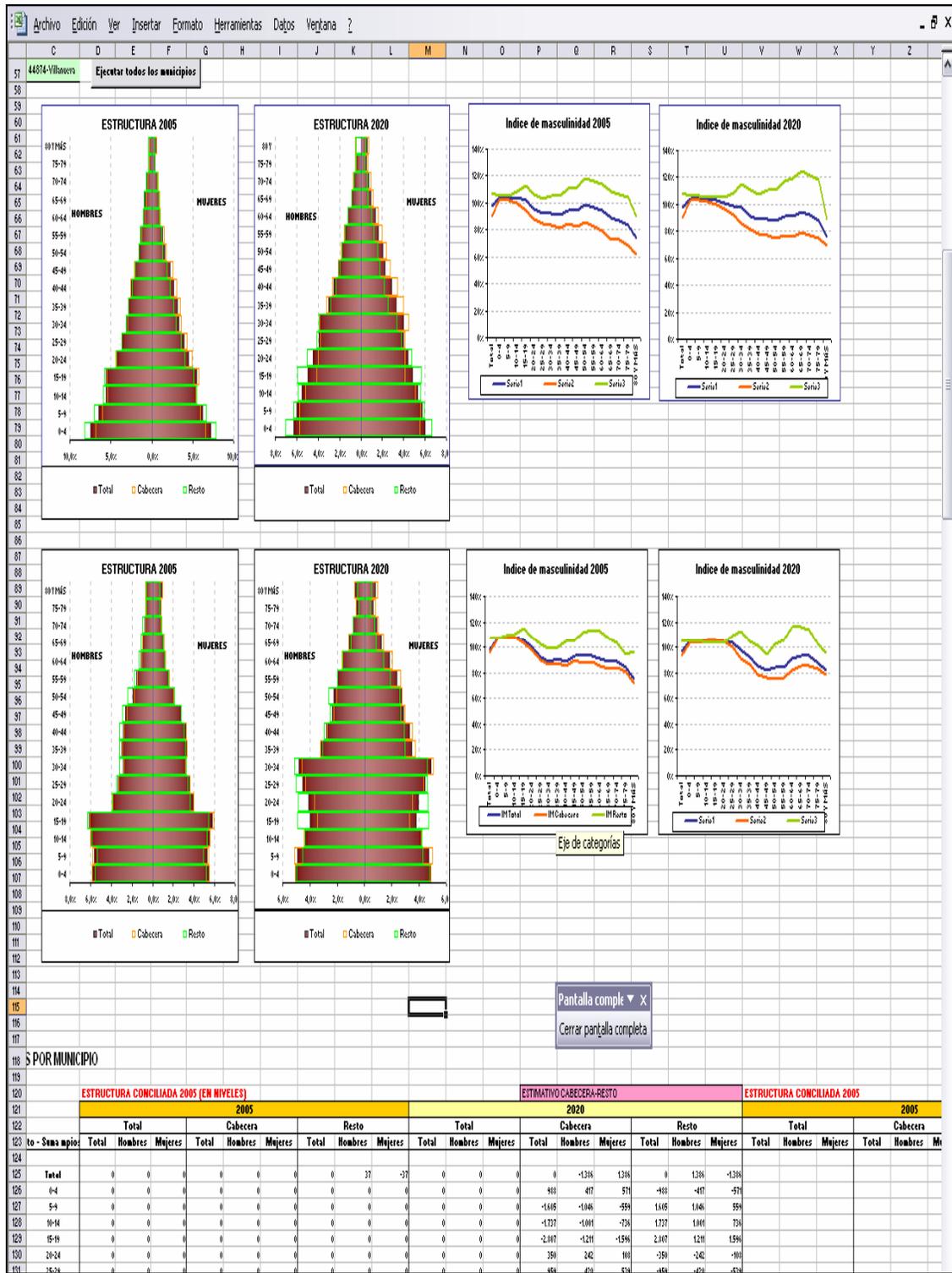
${}_5DP_x^{2020}$ Es el grupo de edad x de la cabecera departamental del 2020

${}_5DP_x^{2005}$ Es el grupo de edad x de la cabecera departamental del 2005

Una vez calculadas estas estructuras se calibran para que la suma de los municipios por grupos de edad y sexo coincida con la estructura departamental del área respectiva.

Para este proceso se contó con la plantilla **Proyecciones Cabecera-Resto 2005-2020 - Estructura Teórica 2020.xls**, la cual efectúa de manera automática los cálculos descritos para cada municipio y departamento.

Figura 4. Vista de la hoja etapa 2 de la plantilla Proyecciones Cabecera-Resto 2005-2020 - Estructura Teórica 2020



c. **Interpolación de las estructuras cabecera resto 2006-2019.** Una vez finalizadas las sub etapas anteriores, tenemos estructuras cabecera resto 2005 y 2020, quedando faltantes los datos para el 2006 hasta el 2019. Este problema se soluciona mediante interpolación lineal por mínimos cuadrados ordinarios - MCO.

La forma lineal generalmente se representa mediante la ecuación $a+bx$ en donde a es la línea que interseca el eje en el plano cartesiano y b es la pendiente de la recta.

Tanto a como b se calculan a través de las ecuaciones de regresión por mínimos cuadrados ordinarios.

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad y$$

$$a = \bar{y} - b\bar{x} \quad \text{En donde:}$$

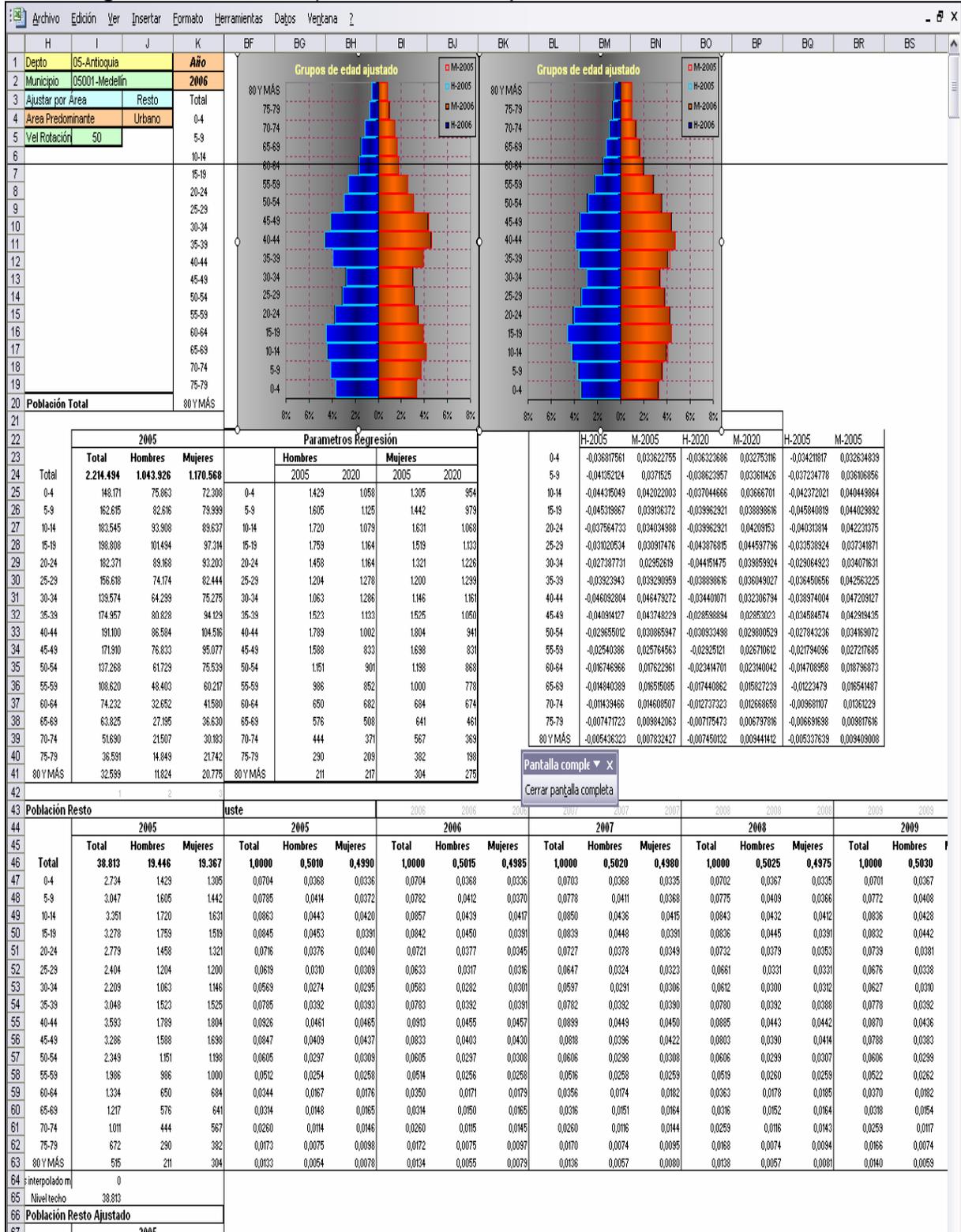
x = variable independiente (se tomaron los años a interpolar)

y = variable dependiente (son los grupos de edad x para el año t ${}_5M N_x^t$)

\bar{x} y \bar{y} son las medias correspondientes de x é y

En esta etapa, se utilizó la plantilla **modelo ajuste MCO.xls**, en donde se puede de manera automática hacer la interpolación con la metodología atrás descrita, para cada municipio por separado en todos los departamentos.

Figura 5. Vista de la plantilla modelo ajuste MCO.xls



Una vez terminada la interpolación se procedió a garantizar que la suma por edades y sexo coincidiera con los respectivos valores departamentales, en caso que esto no sucediera, se distribuían las diferencias proporcionalmente al tamaño del municipio dentro del total departamental

IV- ESTIMACION DE LA POBLACIÓN POR ÁREA, EDAD Y SEXO, PARA LOS DOMINIOS DE LA GEIH, 2000 – 2005.

El objetivo fundamental de este proceso es obtener las estimaciones que requiere de población, la Gran Encuesta Integrada de Hogares – GEIH, para realizar los procesos de expansión teniendo en cuenta las nuevas tendencias determinadas por los resultados del censo 2005.

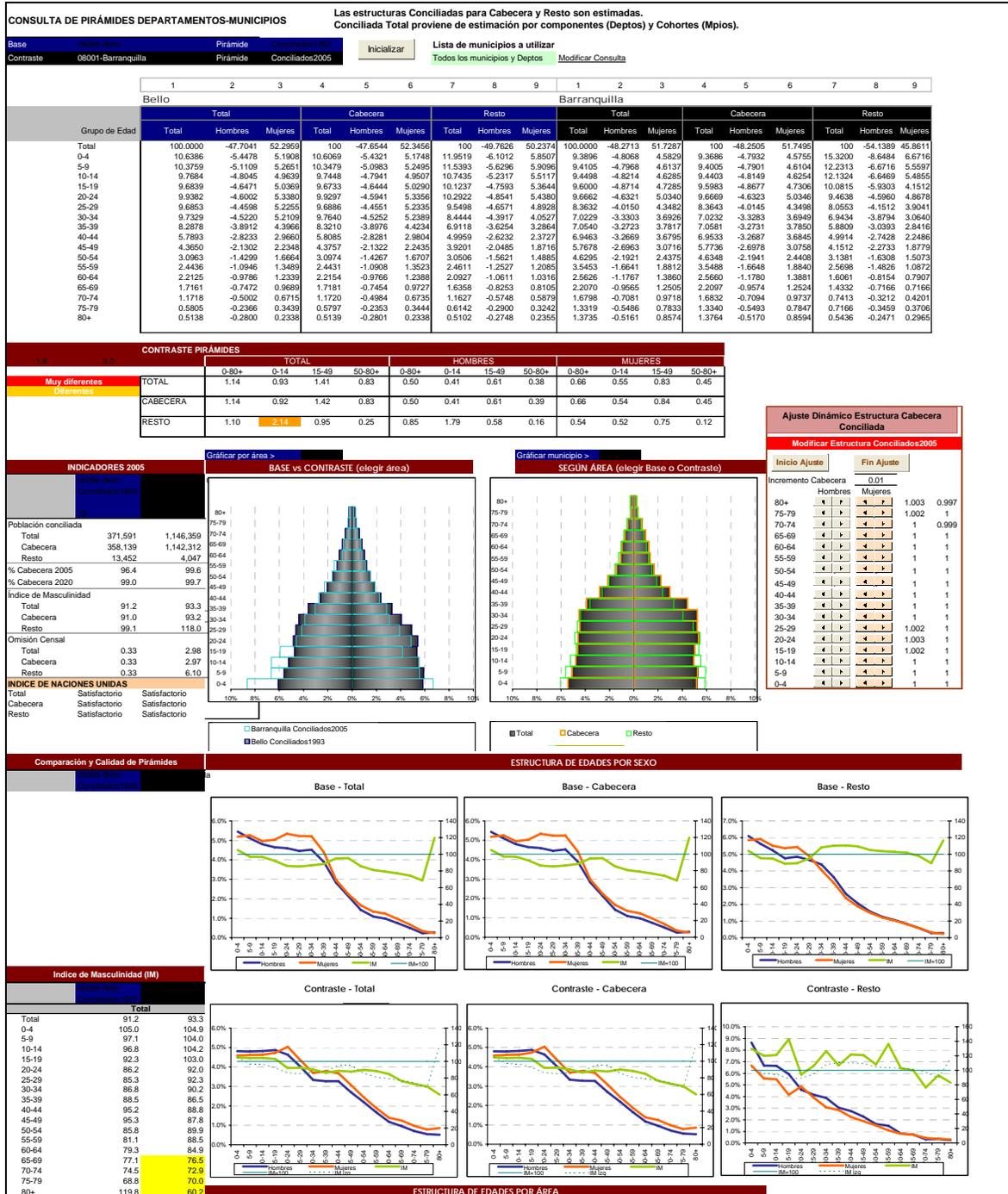
Este proceso, requirió la implementación de tres fases fundamentales. La primera consistió en realizar la calibración del pivote correspondiente al censo de 1993 a nivel cabecera – resto; la segunda, la realización de una interpolación lineal por mínimos cuadrados ordinarios MCO y la tercera, la quincenalización del periodo 2000 – 2005, para cada uno de los dominios de la encuesta.

Fase1: Calibración estructuras cabecera – resto censo 1993. Para realizar este proceso se realiza una homologación de cada municipio con el año 2005 considerando que en el transcurso del periodo intercensal se han creado varios municipios, lo cual requiere garantizar la comparación tanto de niveles como de estructura en el área resto del municipio segregante.

Una vez definida esta homologación entre la geografía de 2005 con la homologada 1993, se realiza el proceso de calibración de la estructura de las áreas cabecera - resto de cada uno de los municipios, con excepción del grupo amazonas que no forma parte del dominio de la encuesta, para lo cual se controlan indicadores como el Índice de Masculinidad por Edad, Índice de Naciones Unidas como Proxy de ajuste de la estructura por edad y sexo, distribución cabecera – resto y la coherencia con la estructura del año 2005

Para realizar este proceso se utiliza la plantilla que aparece en la figura 6, la cual permite realizar la calibración utilizando factores dinámicos de ajuste por grupo de edad y sexo, que en el caso de los dominios de la GEIH, al ajustar el municipio automáticamente realiza un ajuste a la estructura residual departamental. Esta estructura residual se obtiene como la diferencia entre los municipios que pertenecen al dominio en un departamento y la estructura departamental total obtenida por componentes o cohortes según sea el caso del tipo de área.

Figura 5. Vista de la plantilla Estimación Estructuras Conciliadas Suavizado Retro GEIH v2.xls



Fase2: Interpolación Lineal de las Estructuras por Mínimos Cuadrados Ordinarios para cada año del periodo 1993 – 2005. Para realizar este proceso se utilizó el mismo procedimiento y plantilla implementada para obtener las estructuras cabecera – resto del periodo 2006 – 2019, el cual esta referido en el punto c del capítulo anterior.

Resultado de este proceso, se obtiene entonces las estructuras por edad y sexo para cada uno de los dominios de la GEIH, las cuales se aplican a los totales cabecera – resto de cada municipio obtenidos en los pasos iniciales del proceso de proyección³.

Fase3: Interpolación no Lineal de las Estructuras para cada quincena del periodo 2000 – 2009.

Dada las características de la Encuesta Continua de Hogares, a través de la cual se estiman los principales indicadores del mercado laboral, las estimaciones de población por sexo, edad y área se requieren para cada una de las 24 quincenas de año calendario en todo el periodo. Un aspecto importante a tener en cuenta al realizar este proceso, es que al pasar de la quincena 24 del año j a la primera o la 26 del año $(j+1)$ se puede registrar “saltos” que determinan alteraciones en las tendencias de la serie en cada grupo de edad o cohorte.

Con el fin de minimizar éste problema de saltos al cambio de año calendario, se optó por realizar una interpolación de todo el periodo 2000 – 2009 mediante una *Función Splines* la cual está formada por varios polinomios, cada uno definido sobre un subintervalo, que se unen entre sí obedeciendo a ciertas condiciones de continuidad, y cuyos pivotes son las estimaciones anualizadas obtenidas en la fase anterior. La característica de las funciones que determinan el comportamiento de las series por sexo y edad, así como los diferenciales de áreas determinan funciones difíciles de manejar por lo cual interesa encontrar una más sencilla que corresponda a la serie de datos conocida.

Una de las clases de funciones más útiles y mejor conocidas es la de los polinomios algebraicos, es decir el conjunto de funciones de la forma $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 x^0$, donde n es un entero no negativo y $a_n, a_{n-1}, \dots, a_1, a_0$ son constantes reales. La interpolación polinómica se basa en la sustitución de una función o de un conjunto de valores por un polinomio que toma dichos valores. Cuando el número de puntos aumenta, también aumenta el grado del polinomio, que se hace más oscilante (lo cual se traduce en un aumento de los errores). Un enfoque alternativo para conseguir mejores aproximaciones polinomiales es el uso de polinomios de grado menor en subintervalos. Esta es la base de la interpolación con *Splines*. La aproximación polinómica por segmentos

³ Los totales cabecera – resto para el periodo 1993 – 2020 se obtiene mediante un modelo logístico.

más común utiliza polinomios de grado tres entre cada par de puntos consecutivos y recibe el nombre de *Interpolación de Trazadores Cúbicos* o *Splines Cúbicos*.

El modelo matemático que permite estructurar este procedimiento para quincenalizar las series de población por sexo y edad en el periodo 2000 – 2009 esta fundamentado en lo siguiente:

Dados $n+1$ puntos $a = x_0 < x_1 < \dots < x_n = b$, y dada una función f definida en $[a, b]$, una *Función Spline Cúbica* S que interpola la función f , es una función que cumple con las siguientes condiciones:

1. $s(x)$ es un polinomio cúbico denotado por $q_i(x)$ en el intervalo $[x_i, x_{i+1}]$, así:

$$s(x) = q_i(x) \text{ en } [x_i, x_{i+1}], \quad i = 0, 1, \dots, n-1.$$

2. Para que $s(x)$ interpole en los puntos P_0, P_1, \dots, P_n , los $q_k(x)$ han de verificar:

$$\begin{cases} q_i(x_i) = y_i \\ q_i(x_{i+1}) = y_{i+1} \end{cases} \quad i = 0, 1, \dots, n-1. \quad (1)$$

3. En los puntos en común de cada sub-intervalo se cumple

- a. $q_{i-1}(x_i) = q_i(x_i)$ para $i = 1, \dots, n-1$ (lo que asegura la continuidad)
- b. Los polinomios $q_i(x)$ tienen la misma pendiente y la misma concavidad en los nodos que los unen, es decir

$$\begin{cases} q'_{i-1}(x_i) = q'_i(x_i) \\ q''_{i-1}(x_{i+1}) = q''_i(x_{i+1}) \end{cases} \quad i = 1, \dots, n-1. \quad (2)$$

Si $s(x)$ es cúbica a trozos en el intervalo $[x_0, x_n]$, su derivada segunda $s''(x)$ es lineal en el mismo intervalo e interpola en los puntos $(x_i, s''(x_i))$ y $(x_{i+1}, s''(x_{i+1}))$ en $[x_i, x_{i+1}]$. Por tanto, $q_i(x)$ es un polinomio de grado uno que interpola en los puntos $(x_i, s''(x_i))$ y $(x_{i+1}, s''(x_{i+1}))$, y esta dada por:

$$q_i''(x) = \frac{s''(x_i)(x_{i+1} - x) + s''(x_{i+1})(x - x_i)}{x_{i+1} - x_i}, \quad i = 0, 1, \dots, n-1.$$

Denotando con $\alpha_i = s''(x_i)$, $\alpha_{i+1} = s''(x_{i+1})$, $i = 0, \dots, n$,

y

$$h_i = x_{i+1} - x_i, \quad i = 0, \dots, n-1,$$

tenemos

$$q_i''(x) = \frac{\alpha_i(x_{i+1} - x)}{h_i} + \frac{\alpha_{i+1}(x - x_i)}{h_i}, \quad i = 0, 1, \dots, n-1, \quad (3)$$

donde h_i y α_i son constantes. Integrando dos veces:

$$q_i(x) = \frac{\alpha_i(x_{i+1} - x)^3}{6h_i} + \frac{\alpha_{i+1}(x - x_i)^3}{6h_i} + C_i x + D_i, \quad i = 0, 1, \dots, n-1, \quad (4)$$

Donde el término lineal se puede escribir como:

$$C_i x + D_i = A_i(x - x_i) + B_i(x_{i+1} - x),$$

Donde A_i y B_i son constantes arbitrarias, luego tenemos

$$q_i(x) = \frac{\alpha_i(x_{i+1} - x)^3}{6h_i} + \frac{\alpha_{i+1}(x - x_i)^3}{6h_i} + A_i(x - x_i) + B_i(x_{i+1} - x), \quad i = 0, 1, \dots, n-1, \quad (5)$$

Aplicando a (5) las condiciones (1), obtenemos:

$$y_i = \frac{\alpha_i(x_{i+1} - x_i)^3}{6h_i} + B_i(x_{i+1} - x_i) = \frac{\alpha_i h_i^2}{6} + B_i h_i \quad (6)$$

$$y_{i+1} = \frac{\alpha_{i+1}(x_{i+1} - x_i)^3}{6h_i} + A_i(x_{i+1} - x_i) = \frac{\alpha_{i+1} h_i^2}{6} + A_i h_i \quad (7)$$

Despejando de (6) y (7) A_i y B_i , obtenemos

$$A_i = \frac{y_{i+1}}{h_i} - \frac{\alpha_{i+1} h_i}{6}$$

$$B_i = \frac{y_i}{h_i} - \frac{\alpha_i h_i}{6}$$

Al reemplazar A_i y B_i en (5) tenemos la ecuación del Spline $q_i(x)$:

$$q_i(x) = \frac{\alpha_i (x_{i+1} - x)^3}{6h_i} + \frac{\alpha_{i+1} (x - x_i)^3}{6h_i} + \frac{y_{i+1} (x - x_i)}{h_i} - \frac{\alpha_{i+1} h_i (x - x_i)}{6} \\ + \frac{y_i (x_{i+1} - x)}{h_i} - \frac{\alpha_i h_i (x_{i+1} - x)}{6}, \quad i = 0, 1, \dots, n-1. \quad (8)$$

Derivando en (8) obtenemos:

$$q'_i(x) = \frac{-\alpha_i (x_{i+1} - x)^2 + \alpha_{i+1} (x - x_i)^2}{2h_i} + \frac{y_{i+1} - y_i}{h_i} + \frac{\alpha_i - \alpha_{i+1}}{6} h_i \quad (9)$$

Aplicando a (9) las condiciones (2), obtenemos:

$$q'_i(x_i) = \frac{y_{i+1} - y_i}{h_i} - \frac{\alpha_i h_i}{3} - \frac{\alpha_{i+1}}{6} h_i \quad (10)$$

$$q'_i(x_{i+1}) = \frac{y_{i+1} - y_i}{h_i} + \frac{\alpha_{i+1}}{3} h_i + \frac{\alpha_i}{6} h_i \quad (11)$$

Utilizando la condición de continuidad $q'_{i-1}(x_i) = q'_i(x_i)$, se reemplaza i por $i-1$ en (11) para obtener $q'_{i-1}(x_i)$ e igualando a (10) se obtiene:

$$h_{i-1} \alpha_{i-1} + 2(h_{i-1} + h_i) \alpha_i + h_i \alpha_{i+1} = 6 \left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \right), \quad i = 1, \dots, n-1.$$

La ecuación anterior, genera un sistema de $n-1$ ecuaciones lineales con $n+1$ incógnitas $\alpha_0, \alpha_1, \dots, \alpha_n$, lo cual produce un sistema subdeterminado que tiene infinitas soluciones.

Existen varias estrategias para eliminar α_0 de la primera ecuación y α_n de la $(n-1)$ -ésima ecuación, produciendo un sistema tridiagonal de orden $(n-1)$ en las variables $\alpha_1, \alpha_2, \dots, \alpha_{n-1}$, que se puede resolver mediante eliminación Gaussiana sin pivoteo.

Las siguientes alternativas permiten resolver el anterior sistema de ecuaciones

- a. $\alpha_0 = s''(x_0)$ y $\alpha_n = s''(x_n)$. Si se supone que $\alpha_0 = 0$ y $\alpha_n = 0$ se denomina *Spline Natural*.

o bien

- b. $s''(x_0) = f'(x_0)$ y $s''(x_n) = f'(x_n)$, se denomina *Spline Cúbico*.

Para el caso de la obtención de las series de población 2000 – 2007 por sexo y edad para cada uno de los dominios de la Encuesta, se optó por la alternativa *Spline Natural*, que garantiza un mayor suavizamiento de la serie, con lo cual se logra minimizar la posibilidad de “saltos” en la serie que pueden ocasionar distorsiones en las estimaciones de los indicadores.

La implementación de este modelo de interpolación se realizó utilizando el Software libre R versión 2.