II. PREPARACION DE LAS TABLAS MODELO DE MORTALIDAD

Tras experimentar con distintos procedimientos, se optó por una variación del análisis clásico de los componentes principales como modelo analítico. En este procedimiento, las pautas de la mortalidad según la edad reflejadas en el juego de datos básicos refinados se estratificaron en conglomerados mediante métodos gráficos y estadísticos, teniendo cada conglomerado una pauta media distinta de mortalidad según la edad. Luego se ajustó un modelo de componentes principales a las desviaciones de cada pauta de mortalidad según la edad respecto del promedio de su propio conglomerado. La pauta de mortalidad según la edad correspondiente a cada tabla básica de mortalidad fue operacionalizada como vector de los valores de la logit $[{}_nq_x]$, la media del conglomerado como promedios simples de los valores de la logit $[{}_nq_x]$ dentro del conglomerado y las desviaciones de cada pauta respecto del promedio de su conglomerado como las diferencias aritméticas para cada grupo de edades. En todos los casos los grupos de edades fueron los de 0-1, 1-4, 5-9, 10-14, . . . , 80-84.

Los detalles de la preparación de las tablas modelo de mortalidad son los siguientes. Primero, se elaboraron perfiles de las pautas de mortalidad según la edad para cada tabla de mortalidad, mediante dos procedimientos estadísticos y uno gráfico. Los procedimientos estadísticos fueron el de la construcción del perfil linealmente óptima (basada en los vectores propios segundo y tercero) y el del análisis de la conglomeración dinámica (encadenamiento máximo, ultrametría baja)[6]. El procedimiento gráfico fue muy sencillo. Para cada tabla básica de mortalidad se calcularon las razones $R(x) = {}_nq_x/{}_nq_x^w$, en que ${}_nq_x$ es la tasa de mortalidad a la edad x para una tabla de mortalidad dada y ${}_nq_x^w$ es la tasa de mortalidad a la edad x en la tabla modelo de mortalidad de Coale y Demeny para la región occidental, con la misma esperanza de vida a los 10 años de edad.

Los valores de $R(x)$ luego fueron trazados sobre la base de la edad para cada tabla de mortalidad, y los puntos del trazado fueron arreglados ocularmente según la similitud de las pautas. Los tres procedimientos resultaron esencialmente en conglomerados iguales. Había cuatro grupos de pauta clara y unas pocas tablas de mortalidad que en conjunto no se acoplaban bien o fácilmente con ningún otro grupo. Los cuatro grupos o conglomerados son los siguientes: el primero incluye las tablas de mortalidad de los países latinoamericanos de Colombia, Costa Rica, El Salvador, Guatemala, Honduras, México y Perú, así como los países no americanos de Filipinas, Sri Lanka y Tailandia. El segundo conglomerado corresponde a la pauta muy característica de las tablas de mortalidad de Chile. El tercero comprende las tablas de la India, Irán, la región de Matlab, de Bangladesh, y Túnez. El cuarto conglomerado comprende las tablas de Guyana, Hong Kong, República de Corea, Singapur y Trinidad y Tabago para los varones, y Guyana, Singapur y Trinidad y Tabago para las mujeres. Las cuatro pautas han sido denominadas pauta latinoamericana, pauta chilena, pauta sudasiática y pauta del Lejano Oriente, conforme a la región geográfica que predomina dentro de cada grupo. Las tablas de mortalidad de Israel y Kuwait así como las de las mujeres de Hong Kong y la República de Corea no se acoplaban con ninguno de los conglomerados y por ello fueron omitidas del análisis de los componentes principales y sólo se incluyeron en la elaboración de la pauta general de la mortalidad que se describe más abajo.

Dentro de cada conglomerado se calcularon los valores de ${}_nD_x^{ij}$, que para cada grupo de edades $(x, x+n)$ se definen como la diferencia entre los valores de la logit $[{}_nq_x]$ para la tabla de mortalidad j del conglomerado i y la media de los

[6] Estos procedimientos se explican en J. A. Hartigan, *Clustering Algorithms* (Nueva York, John Wiley and Sons, 1975), y en P. H. Sneath y R. R. Sokal, *Numerical Taxonomy—The Principles and Practices of Numerical Classification* (San Francisco, Calif., W. H. Freeman, 1973).

CUADRO 3. CORRELACIONES ENTRE VALORES DE ${}_nD_x^{ij}$ CALCULADOS A BASE DE TABLAS DE MORTALIDAD DE LOS VARONES

	Edad x																	
Edad x	*0*	*1*	*5*	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*	*70*	*75*	*80*
0		0.93	0.79	0.82	0.79	0.83	0.84	0.83	0.85	0.83	0.82	0.79	0.77	0.75	0.74	0.68	0.59	0.35
1			0.87	0.85	0.78	0.82	0.84	0.84	0.86	0.85	0.84	0.82	0.81	0.81	0.79	0.72	0.63	0.42
5				0.96	0.89	0.88	0.88	0.89	0.91	0.92	0.93	0.92	0.93	0.92	0.91	0.87	0.85	0.69
10					0.97	0.95	0.93	0.92	0.93	0.94	0.95	0.93	0.94	0.92	0.93	0.90	0.87	0.68
15						0.97	0.95	0.92	0.93	0.93	0.94	0.93	0.93	0.90	0.91	0.87	0.85	0.65
20							0.98	0.96	0.96	0.96	0.95	0.93	0.93	0.89	0.89	0.82	0.79	0.58
25								0.99	0.98	0.98	0.96	0.94	0.94	0.89	0.89	0.80	0.79	0.55
30									0.98	0.98	0.96	0.94	0.93	0.89	0.87	0.77	0.76	0.52
35										0.99	0.99	0.96	0.96	0.91	0.90	0.82	0.80	0.56
40											0.99	0.98	0.97	0.93	0.91	0.84	0.81	0.58
45												0.99	0.98	0.95	0.94	0.87	0.84	0.63
50													0.99	0.98	0.95	0.89	0.86	0.66
55														0.98	0.97	0.92	0.90	0.69
60															0.96	0.93	0.89	0.72
65																0.95	0.93	0.73
70																	0.97	0.83
75																		0.89
80																		

CUADRO 4. CORRELACIONES ENTRE VALORES DE $_nD_x^{ij}$ CALCULADOS A BASE DE TABLAS DE MORTALIDAD DE LAS MUJERES

	Edad x																
Edad x	*0*			*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*	*70*	*75*	*80*
0	0.90	0.73	0.80	0.80	0.77	0.78	0.81	0.83	0.84	0.83	0.80	0.76	0.75	0.68	0.59	0.54	0.25
1		0.89	0.89	0.91	0.90	0.91	0.92	0.92	0.91	0.89	0.85	0.84	0.82	0.79	0.72	0.66	0.40
5			0.94	0.94	0.94	0.94	0.93	0.93	0.92	0.92	0.91	0.92	0.91	0.91	0.88	0.82	0.62
10				0.96	0.94	0.94	0.94	0.95	0.94	0.93	0.91	0.91	0.90	0.89	0.87	0.84	0.63
15					0.99	0.98	0.98	0.98	0.96	0.95	0.92	0.91	0.88	0.87	0.86	0.83	0.65
20						0.99	0.99	0.98	0.95	0.94	0.91	0.90	0.87	0.88	0.87	0.84	0.67
25							0.99	0.99	0.96	0.95	0.92	0.92	0.89	0.89	0.87	0.85	0.68
30								1.00	0.98	0.97	0.94	0.93	0.90	0.89	0.87	0.84	0.65
35									0.99	0.97	0.95	0.93	0.91	0.90	0.87	0.84	0.64
40										0.99	0.98	0.96	0.95	0.91	0.86	0.83	0.61
45											0.99	0.98	0.97	0.93	0.88	0.85	0.62
50												0.99	0.98	0.94	0.88	0.83	0.61
55													0.99	0.97	0.91	0.86	0.63
60														0.97	0.91	0.86	0.61
65															0.97	0.93	0.71
70																0.98	0.83
75																	0.89
80																	

valores correspondientes a todas las tablas de mortalidad dentro del conglomerado i, en que

$$\text{logit } [_nq_x] = 1/2\, ln\left(\frac{_nq_x}{1 - {_nq_x}}\right).$$

Como era de esperar, un valor de $_nD_x^{ij}$ de un grupo de edades guarda una elevada correlación con los valores de otros grupos de edades. La matriz de correlación se presenta en los cuadros 3 y 4 para los varones y las mujeres. Los coeficientes de correlación están generalmente por encima de 0,80, y las correlaciones más bajas se observan principalmente en los grupos de edad más avanzada. Dado que cada conglomerado está compuesto por un juego congruente de tablas de mortalidad con similares pautas de mortalidad según la edad, el vector $_nD_x^{ij}$ de cada tabla básica de mortalidad puede considerarse como una indicación de la pauta del cambio de la mortalidad según la edad, es decir que indica cómo varía la mortalidad en cada edad. Partiendo de la premisa de que la pauta de mortalidad según la edad es invariante respecto de la pauta del conglomerado[7], podemos expresar la estructura de la mortalidad por edades de cualquier país (definida por sus valores de $[_nq_x]$ como $_nY_x^{ij} = {_n\bar{Y}_x^i} + a_{1j}U_{1x}$, en que $_nY_x^{ij}$ es igual a la logit de la función $_nq_x$ de la tabla de mortalidad j del conglomerado i, y $_n\bar{Y}_x^i$ es igual a la media de $_nY_x^{ij}$ dentro de cada conglomerado. El vector designa entonces en promedio la pauta del cambio de mortalidad según la edad (una especie de promedio de los valores de $_nD_x^{ij}$) y a_{1j} designa la magnitud del cambio. Esto es esencialmente un modelo de un componente, y el vector U_{1x}, llamado vector del primer componente principal, indica la pauta del cambio de la mortalidad, y su coeficiente (a_{1j}), llamado factor de carga, indica la magnitud del cambio correspondiente a la tabla de mortalidad j.

Por supuesto, este modelo componente no explicará todas las variaciones de las estructuras de la mortalidad según la edad que se observan en las tablas de mortalidad del conjunto de datos refinados. Es posible calcular nuevos grupos de desviaciones como la diferencia entre la expresión de los valores empíricos de la expresión $[_nq_x]$ y los pronosticados a base del modelo de un componente. Si optamos por que U_{2x} designe una especie de pauta media según la edad de estas desviaciones de segundo grado, y por que a_{2j} designe la magnitud de esta pauta de desviaciones respecto de una tabla de mortalidad j, podremos formular un modelo de dos componentes como

$$_nY_x^{ij} \doteq {_n\bar{Y}_x^i} + \sum_{m=1}^{2} a_{mj}U_{mx}.$$

De la misma manera pueden calcularse modelos de tres, cuatro y hasta 18 componentes. En consecuencia, la forma funcional del modelo puede expresarse como

$$_nD_x^{ij} = {_nY_x^{ij}} - {_n\bar{Y}_x^i} \doteq \sum_{m=1}^{k} a_{mj}U_{mx} \qquad (1)$$

en que: $_nY_x^{ij}$ es igual a la logit de la función $_nq_x$ (probabilidad de fallecer entre las edades de x y $x + n$) para la tabla de mortalidad j del conglomerado i; $_n\bar{Y}_x^i$ es igual a la media de $_nY_x^{ij}$ dentro de cada conglomerado; a_{mj} es igual a la carga factorial del vector de componente principal m para el país j en el análisis de componentes principales; u_{mx} es igual al elemento del vector de componente principal m correspondiente al grupo de edades $(x, f + n)$; y k es el número de componentes principales.

Para los fines de la aplicación suele ser más conveniente expresar el modelo como

$$_nY_x^{ij} \doteq {_n\bar{Y}_x^i} + \sum_{m=1}^{k} a_{mj}U_{mx}. \qquad (2)$$

Cuando k es igual a 1, el modelo se llama modelo de un componente; cuando k es igual a 2, modelo de dos componentes, y así sucesivamente. El modelo de componentes principales es similar a los procedimientos acostumbrados de regresión lineal, puesto que se pueden averiguar valores de parámetros que minimizan las sumas de desviaciones cuadráticas. En este caso se averiguan los vectores U_{1x}, U_{2x}, U_{3x}, . . . , U_{kx}, que minimizan en secuencia la suma de desviaciones cuadráticas entre los valores de $_nD_x^{ij}$ reales y pronosticados. Las distancias entre los valores reales y los pronosticados se miden como distancias perpendiculares

[7] La hipótesis de invariancia de la pauta del cambio en la mortalidad según la edad respecto de la pauta del conglomerado parece muy firme a primera vista. Sin embargo, la aplicación por separado del análisis de componentes principales a la pauta latinoamericana, la pauta chilena y la pauta del Lejano Oriente indicó vectores de un componente muy similares dentro de cada conglomerado. (Dado que entre las tablas de mortalidad incluidas en la pauta sudasiática había poca variación en los niveles de mortalidad, resultó imposible efectuar un análisis por separado de componentes principales para ese conglomerado). Esta comprobación empírica permitió superponer una sola pauta de cambio de la mortalidad según la edad a las cuatro pautas básicas.

CUADRO 5. PAUTA MEDIA DE MORTALIDAD PARA CADA CONGLOMERADO DEFINIDA POR LOS VALORES DE LA LOGIT [$_nq_x$]

	Conglomerado de varones						*Conglomerado de mujeres*				
Edad x	*Latinoamericana*	*Chilena*	*Sudasiática*	*Del Lejano Oriente*	*General*	*Edad x*	*Latinoamericana*	*Chilena*	*Sudasiática*	*Del Lejano Oriente*	*General*
0	−1.12977	−1.04722	−0.97864	−1.53473	−1.27638	0	−1.22452	−1.12557	−0.97055	−1.42596	−1.35963
1	−1.49127	−1.81992	−1.24228	−2.15035	−1.78957	1	−1.45667	−1.82378	−1.15424	−1.95200	−1.77385
5	−2.13005	−2.42430	−2.01695	−2.61442	−2.35607	5	−2.13881	−2.52319	−1.93962	−2.55653	−2.39574
10	−2.40748	−2.52487	−2.44280	−2.66392	−2.55527	10	−2.46676	−2.63933	−2.36857	−2.68018	−2.64549
15	−2.21892	−2.24491	−2.35424	−2.42326	−2.34263	15	−2.31810	−2.38847	−2.19082	−2.33095	−2.44766
20	−2.01157	−2.02821	−2.27012	−2.23095	−2.16193	20	−2.14505	−2.20417	−2.09358	−2.15952	−2.28991
25	−1.93591	−1.90923	−2.16833	−2.15279	−2.09109	25	−2.03883	−2.09701	−2.04788	−2.03377	−2.18850
30	−1.86961	−1.78646	−2.05942	−2.05765	−2.00215	30	−1.93924	−1.99128	−1.95922	−1.94554	−2.08535
35	−1.76133	−1.66679	−1.90053	−1.89129	−1.86781	35	−1.83147	−1.87930	−1.87311	−1.82299	−1.97231
40	−1.64220	−1.52497	−1.71213	−1.68244	−1.70806	40	−1.74288	−1.75744	−1.76095	−1.69084	−1.84731
45	−1.49651	−1.37807	−1.51120	−1.47626	−1.52834	45	−1.62385	−1.61558	−1.61425	−1.52189	−1.69291
50	−1.34160	−1.21929	−1.28493	−1.23020	−1.33100	50	−1.47924	−1.45886	−1.39012	−1.33505	−1.50842
55	−1.15720	−1.03819	−1.08192	−1.02801	−1.12934	55	−1.28721	−1.26115	−1.15515	−1.13791	−1.30344
60	−0.96945	−0.84156	−0.84671	−0.77148	−0.91064	60	−1.07443	−1.05224	−0.90816	−0.93765	−1.08323
65	−0.74708	−0.63201	−0.62964	−0.54696	−0.68454	65	−0.83152	−0.80346	−0.68011	−0.72718	−0.84402
70	−0.52259	−0.42070	−0.40229	−0.32996	−0.45685	70	−0.59239	−0.58202	−0.43231	−0.50916	−0.59485
75	−0.29449	−0.21110	−0.19622	−0.11911	−0.23002	75	−0.35970	−0.35093	−0.17489	−0.28389	−0.34158
80	−0.04031	0.01163	−0.00129	0.10572	0.00844	80	−0.08623	−0.10587	0.05948	−0.01285	−0.06493

(ortogonales), y no como distancias verticales. Puede demostrarse que los vectores U_{mx} son simplemente los vectores propios de la matriz de covariancias de los valores de ${}_nD_x^{ij}$ [8]. Aun cuando hace falta un número de componentes igual al de variables (grupos de edades) para explicar la totalidad de la variación en los valores de ${}_nD_x^{ij}$ (en el presente caso hacen falta 18 componentes porque hay 18 grupos de edades), a menudo los primeros componentes reflejan una magnitud de variación suficiente para aplicarla a diversas finalidades. En el caso de la tabla modelo de mortalidad un solo componente explicó alrededor del 90% de la variación, y tres componentes explicaron el 97%.

En el cuadro 5 se indica, por sexos, la pauta media de mortalidad para cada conglomerado operacionalizado por la media de los valores de $[{}_nq_x]$ de las tablas de mortalidad que incluye. También se indica una "pauta general", que se calculó promediando los valores de la $[{}_nq_x]$ de todas las tablas de mortalidad en el conjunto de los datos refinados sin tomar en consideración ningún conglomerado. En el capítulo III *infra* se describen detalladamente las características de las diversas pautas.

En el cuadro 6 se presentan los tres vectores de componentes principales por sexos. Como se esperaba, el primer componente principal da el modelo de la pauta del cambio en la mortalidad según la edad. A juzgar por este componente, a medida que la mortalidad declina el cambio se hace mayor en los años de la niñez y se reduce con el aumento de la edad. Las declinaciones durante la infancia son algo menores que durante la niñez, como las que ocurren en los últimos años de la edad mediana. La expresión "pauta del cambio de la mortalidad" utilizada en este contexto se refiere, por supuesto, al cambio en la función de la logit $[{}_nq_x]$ de la tabla de mortalidad. Dado que, salvo cuando los valores ${}_nq_x$ son bastante altos, la logit $[{}_nq_x]$ está muy cerca de

[8] Para una exposición más rigurosa del análisis de los componentes principales véase, por ejemplo, D. F. Morrison, *Multivariate Statistical Method* (Nueva York, McGraw-Hill, 1976).

CUADRO 6. PRIMEROS TRES COMPONENTES PRINCIPALES

	Varones			Mujeres		
Edad x	*1er. componente* U_{1x}	*2do. componente* U_{2x}	*3er. componente* U_{3x}	*1er. componente* U_{1x}	*2do. componente* U_{2x}	*3er. componente* U_{3x}
0	0.23686	−0.46007	0.09331	0.18289	−0.51009	0.23944
1	0.36077	−0.68813	−0.29269	0.31406	−0.52241	−0.11117
5	0.33445	0.06414	−0.47139	0.31716	0.08947	0.07566
10	0.30540	0.12479	−0.17403	0.30941	0.03525	0.06268
15	0.28931	0.24384	0.10715	0.32317	0.03132	−0.26708
20	0.28678	0.10713	0.28842	0.32626	0.07843	−0.39053
25	0.27950	0.06507	0.33620	0.30801	0.06762	−0.28237
30	0.28023	0.03339	0.33692	0.29047	0.00482	−0.14277
35	0.26073	0.02833	0.21354	0.25933	−0.01409	−0.05923
40	0.23626	0.06473	0.15269	0.22187	−0.02178	0.18909
45	0.20794	0.08705	0.06569	0.19241	0.01870	0.24773
50	0.17804	0.10620	0.00045	0.17244	0.04427	0.33679
55	0.15136	0.11305	−0.03731	0.15729	0.08201	0.34121
60	0.13217	0.09467	−0.10636	0.14282	0.08061	0.38290
65	0.12243	0.10809	−0.11214	0.12711	0.15756	0.26731
70	0.11457	0.14738	−0.22258	0.11815	0.24236	0.14442
75	0.10445	0.21037	−0.19631	0.11591	0.30138	0.09697
80	0.08878	0.30918	−0.38123	0.09772	0.50530	−0.13377

CUADRO 7. PROPORCIÓN DE LA VARIACIÓN DE LA MORTALIDAD EXPLICADA POR COMPONENTES Y EDADES

	Varones			Mujeres		
	Proporción de la variación explicada por			*Proporción de la variación explicada por*		
Edad x	*Un componente*	*Dos componentes*	*Tres componentes*	*Un componente*	*Dos componentes*	*Tres componentes*
0	0.772	0.931	0.935	0.687	0.911	0.932
1	0.815	0.977	0.993	0.870	0.971	0.973
5	0.909	0.911	0.965	0.921	0.924	0.925
10	0.951	0.960	0.969	0.935	0.936	0.937
15	0.914	0.950	0.954	0.974	0.974	0.986
20	0.940	0.947	0.975	0.967	0.969	0.994
25	0.946	0.949	0.990	0.974	0.976	0.991
30	0.933	0.934	0.974	0.982	0.982	0.986
35	0.959	0.960	0.979	0.985	0.986	0.986
40	0.962	0.966	0.978	0.962	0.962	0.975
45	0.964	0.974	0.976	0.954	0.954	0.982
50	0.933	0.951	0.951	0.913	0.915	0.976
55	0.934	0.963	0.964	0.898	0.908	0.982
60	0.884	0.909	0.926	0.866	0.878	0.987
65	0.870	0.907	0.929	0.843	0.897	0.962
70	0.763	0.831	0.918	0.789	0.928	0.949
75	0.700	0.854	0.928	0.725	0.931	0.940
80	0.386	0.641	0.854	0.412	0.874	0.887
Todas las edades en conjunto	0.892	0.940	0.967	0.913	0.952	0.968

la mitad de los valores $\ell n\ {}_nq_x$, es posible considerar que los elementos del primer componente representan el cambio proporcional de los valores de ${}_nq_x$.

El segundo componente parece explicar principalmente las diferencias características entre las tablas de mortalidad en la relación de la mortalidad de los menores de 5 años con la mortalidad de los mayores de esa edad, diferencias que no quedaron plenamente explicadas ni por la conglomeración inicial de las pautas de mortalidad en cuatro grupos ni por la pauta del cambio de la mortalidad según la edad descrita por el primer componente.

El tercer componente parece afectar a la mortalidad durante los años de la procreación en las mujeres y durante un número diverso de edades en los hombres.

El juego de tablas modelo de mortalidad del anexo I es un modelo de un componente, basado en las cinco pautas medias (los cuatro grupos de pautas y la pauta general), y en la pauta de cambio en la mortalidad según la edad definida por el primer componente principal. Como se ve en el cuadro 7, semejante modelo de un componente explica la mayor parte de la variación entre todas la tablas básicas de mortalidad. Concretamente, la variación del 89% para los varones y del 91% para las mujeres en los valores de la logit $[{}_nq_x]$ del conjunto de los datos básicos se explica por el primer componente en la primera conglomeración regional. Sin embargo, como se observa claramente en el cuadro 7, la magnitud de variación de la mortalidad explicada no es idéntica para todos los grupos de edades. Entre los varones, el 90% o más de la variación sólo se explica para las edades de 5 a 59 años; entre las mujeres, esta gama se extiende de los 5 a los 54 años. El segundo y el tercer componentes explican en total un 5% y un 3% adicionales respectivamente de variación para los varones, y un 4% y un 2% respectivamente para las mujeres. Para ambos sexos, extendiendo el modelo a dos componentes se explica más del 90% de la variación para todos los grupos de edades, menos unos pocos de las edades más avanzadas.

Los modelos presentados en el anexo I son modelos de un componente. No obstante, puede aprovecharse la disponibilidad del segundo y el tercer componentes y la variación adicional que ellos explican para elaborar otros modelos de pautas según la edad. Estas posibilidades se consideran seguidamente, en el capítulo IV.