Determinating Model Life Table Converged By Urban Mortality in Turkey through Kernel Estimation and Hellinger Distance

Serpil CULA
Associate Professor
Başkent University

Şeref HOŞGOR
Assistant Professor
Başkent University

Abstract:

It is of the utmost importance that the model life table that is to be utilised is determined where a population registration system and a registered mortality data is not available and incomplite. In Turkey, the demographic researchers who study the field of mortality have used various model life tables. In this study, kernel estimation method and Hellinger distance have been used.

Key Words: Model Life Table, Hellinger Distance, Kernel Estimation Method

I. Introduction

In Turkey, the researchers who work on life tables and mortality have had to use the technique of indirect estimation due to the lack of mortality data. The most important phase of undertaking mortality estimations or creating life tables making use of indirect techniques is determining the appropriate life models and applying them to the technique used. In our country, a number of researchers have used different model life tables in spite of the fact that they all utilised the same technique. For example, Hancıoğlu, used the Latin American model, Hoşgör the East and West model and Toros the West model. Yet, all of the researchers agree on the assumption that Turkey fails in none of these models as a whole in this category. Thus, the researchers have made estimations on the assumption that the model they have used represents Turkey.

II. Methodology

In this study, it is aimed to achieve the model life table converged by the mortality that is given in Equation (1) by means of using kernel estimation which is a parametric estimation method and hellinger distance.

$$\hat{p}_i = \frac{\sum_{j=1}^{K} W\left(\frac{i/K - j/K}{h}\right)\bar{p}_j}{\sum_{j=1}^{K} W\left(\frac{i/K - j/K}{h}\right)} . \qquad (1)$$

In Equation (1); K (i=1,2,...,K) is the number of group number; n_i, is the observation number in terms of each group; n is the total number of observation and $\bar{p}_i$ is the relative frequency for each group $(\bar{p}_i = n_i/n)$. h is a positive number, usually called the bandwidth or window width. W kernel function can be assumed to be any function that bears the features of probability function. When W kernel function is taken as a continuous function, it also maintains its features in kernel estimation. The selection of the relevant W kernel estimation must be determined based on the features of time, calculation efficiency and derivativeness (Silverman, 1986). It is known that W kernel functions do not create any significant changes on estimation whereas the bandwidth selection significantly affects the performance of kernel estimation in kernel estimations defined by the mean average which is obtained by means of utilizing weights from kernel estimation with centered observations. The selection of h bandwidth in kernel estimation is very important. When h is chosen to be a very low value, the estimation will be crudely made due to the fact that weighted average process applied to obtain kernel estimation in every point will be based on limited and less number of observations. This kind of estimation is called as less smoothed estimation.

On the other hand, if the value of h is too high, the achieved estimation value can be significantly deviated from its real value due to the fact that weighted average process will be based on observations in quite higher number. This kind of estimation is called as more smoothed estimation. The most appropriate value for h bandwidth can be acquired by utilizing some error criteria.

Rosenblatt (1956) states that additive mean of error squares which is widely used and easily monitored general criterion for the accuracy of density function's kernel estimator is preferred because of its mathematical easiness. The h bandwidth that makes these criteria the lowest is taken as the most proper bandwidth. The most appropriate bandwidth value that makes the collected error squares mean the lowest is dependent on the second order derivative of unknown f value. Many methods have been suggested to achieve h bandwidth (Wand and Jones, 1995; Horova, and Zelinka, 2007). Each developed method has advantages and disadvantages compared to others (Cula, 1998). In this study biased cross-validation has been used suggested by Scott and Terrell (Scott and Terrell, 1987).

In the study, sex and age analysis of 2003 and 2008 death data that was collected from cities and provinces by the Turkish Statistics Institute (TUIK) has been used (TUIK, 2003 and 2008). Before starting this study about the model life table that urban deaths converge in Turkey, death data sex analysis needs to be adjusted. Criteria to be used related to conversion are the rate of death in age groups. The more converges that rate of dispersion to the data ratio of death level chosen from the model life table the more it is recommended to use the found model.
In order to find out the kernel density of death data, the h bandwidth (window) of the data must be found. Common biased cross validation method has been used to find h bandwidth. As result of the analysis done for 2003 h bandwidth is 0,01 for both female and male; and for 2008 it is 0,03 for male data and 0,015 for female data. For W core function standard normal kernel function has been used. Found h bandwidth and standard core function are placed in Equation (1) to find core forecast values and adjusted single age dispersion rate for both sexes for 2003 and 2008 has been calculated.

In the study done by Hoşgör in 2000 it was forecast that the death level would be between 20,91 and 24,00 for female and 20,03 and 22,59 for male (Hosgör, 1997). When middle points of both data are taken in 2000 we reach to 22,45 for female and 21,31 for male. In the study it was agreed to assume 2003 death levels are 23,00 for female and around 22,00 for male because both values are within the forecast limit. In another study done by Hoşgör and Tansel in 2011 death levels were forecasted 23,02 for male and 23,69 for female. In this study those indicated death levels have been used.

Mortality rates of every age group are obtained by converting the number of deaths of age groups into rates, sums of which are equal to 1, at level 23 for female and level 22 for male in 2003 and by using Coale and Demany's (1983) Life Table models (East, West, North, South). Death level found for 2008 for female is 23,69 and 23,02 for male. Similarly, mortality rates of every age group are obtained by converting the number of deaths of age groups into rates, sums of which are equal to 1, at all mortality levels and by using Coale and Demany's (1983) Life Table models (East, West, North, South).

Turkey's 2003 and 2008 urban death's stochastic age distribution and age distribution of mortality rate that was obtained from the model life table's death levels have been standardised and became comparable. Reached results have been shown with their core forecast values on Table 1 for female, on Table 2 for male for 2003 and for 2008 they were shown on Table 3 for female and on Table 4 for male.

III. Findings

Table 1. Interpolation And Obtained By Kernel Method Of Female Deaths Numbers And Rates For 2003

AGE GROUP	WEST DEATH NUMBER	EAST DEATH NUMBER	NORTH DEATH NUMBER	SOUTH DEATH NUMBER	WEST DEATH RATE	EAST DEATH RATE	NORTH DEATH RATE	SOUTH DEATH RATE	2003 KERNEL DEATH RATE
	LEVEL23	LEVEL23	LEVEL23	LEVEL23	LEVEL23	LEVEL23	LEVEL23	LEVEL23	h=0,01
0	1530	2139	1878	4132	0,0153	0,0214	0,0188	0,0413	0,0401
1-4	240	256	414	760	0,0024	0,0026	0,0041	0,0076	0,0318
5-9	127	107	195	125	0,0013	0,0011	0,0020	0,0013	0,0058
10-14	105	94	178	99	0,0011	0,0009	0,0018	0,0010	0,0047
15-19	175	154	297	145	0,0018	0,0015	0,0030	0,0015	0,0081
20-24	250	221	428	215	0,0025	0,0022	0,0043	0,0022	0,0093
25-29	317	266	514	264	0,0032	0,0027	0,0051	0,0026	0,0094
30-34	406	345	567	328	0,0041	0,0035	0,0057	0,0033	0,0111
35-39	568	494	687	415	0,0057	0,0049	0,0069	0,0042	0,0143
40-44	859	746	1035	624	0,0086	0,0075	0,0104	0,0062	0,0198
45-49	1400	1221	1394	903	0,0140	0,0122	0,0139	0,0090	0,0283
50-54	2155	1905	2175	1435	0,0216	0,0191	0,0218	0,0144	0,0403
55-59	3322	2880	2921	2087	0,0332	0,0288	0,0292	0,0209	0,0485
60-64	5177	4702	4625	3410	0,0518	0,0470	0,0463	0,0341	0,0720
65-69	8426	7942	7415	5875	0,0843	0,0794	0,0742	0,0588	0,0974
70-74	12973	13149	11400	10558	0,1297	0,1315	0,1140	0,1056	0,1419
75+	61969	63379	63878	68623	0,6197	0,6338	0,6388	0,6862	0,1570

Table 2. Interpolation And Obtained By Kernel Method Of Male Deaths Numbers And Rates For 2003

AGE GROUP	WEST DEATH NUMBER	EAST DEATH NUMBER	NORTH DEATH NUMBER	SOUTH DEATH NUMBER	WEST DEATH RATE	EAST DEATH RATE	NORTH DEATH RATE	SOUTH DEATH RATE	2003 KERNEL DEATH RATE
	LEVEL22	LEVEL22	LEVEL22	LEVEL22	LEVEL22	LEVEL22	LEVEL22	LEVEL22	h=0,01
0	3099	3889	3174	5633	0,0310	0,0389	0,0317	0,0563	0,0444
1-4	599	537	901	1177	0,0060	0,0054	0,0090	0,0118	0,0344
5-9	332	263	521	257	0,0033	0,0026	0,0052	0,0026	0,0063
10-14	272	251	381	228	0,0027	0,0025	0,0038	0,0023	0,0052
15-19	504	494	770	327	0,0050	0,0049	0,0077	0,0033	0,0097
20-24	699	705	1122	449	0,0070	0,0071	0,0112	0,0045	0,0130
25-29	698	714	1129	494	0,0070	0,0071	0,0113	0,0049	0,0135
30-34	783	781	1207	658	0,0078	0,0078	0,0121	0,0066	0,0141
35-39	1001	975	1322	799	0,0100	0,0098	0,0132	0,0080	0,0191
40-44	1464	1415	1618	1193	0,0146	0,0142	0,0162	0,0119	0,0300
45-49	2317	2279	2126	1803	0,0232	0,0228	0,0213	0,0180	0,0460
50-54	3610	3801	3249	2864	0,0361	0,0380	0,0325	0,0286	0,0670
55-59	5581	5823	4221	4291	0,0558	0,0582	0,0422	0,0429	0,0774
60-64	8103	8266	6310	6223	0,0810	0,0827	0,0631	0,0622	0,0985
65-69	11217	11226	9054	8896	0,1122	0,1123	0,0905	0,0890	0,1154
70-74	14463	14639	12209	13058	0,1446	0,1464	0,1221	0,1306	0,1470
75+	45257	43941	50685	51651	0,4526	0,4395	0,5068	0,5165	0,1200

Table 3. Interpolation And Obtained By Kernel Method Of Female Deaths Numbers And Rates For 2008

AGE GROUP	WEST DEATH NUMBER	EAST DEATH NUMBER	NORTH DEATH NUMBER	SOUTH DEATH NUMBER	WEST DEATH RATE	EAST DEATH RATE	NORTH DEATH RATE	SOUTH DEATH RATE	2008 KERNEL DEATH RATE
	LEVEL2 3,69	LEVEL2 3,69	LEVEL2 3,69	LEVEL2 3,69	LEVEL2 3,69	LEVEL2 3,69	LEVEL2 3,69	LEVEL2 3,69	h=0,015
0	1099	1601	1463	3547	0,010987	0,016015	0,014633	0,035469	0,0401
1-4	145	164	275	558	0,001455	0,001635	0,002753	0,005578	0,0318
5-9	82	71	131	89	0,000821	0,000711	0,001308	0,000891	0,0058
10-14	69	65	128	73	0,000691	0,00065	0,001276	0,000728	0,0047
15-19	117	109	225	107	0,00117	0,001091	0,002245	0,001071	0,0081
20-24	170	158	332	161	0,0017	0,001582	0,003321	0,001612	0,0093
25-29	219	191	402	201	0,00219	0,001908	0,004015	0,002005	0,0094
30-34	287	253	441	254	0,002866	0,002532	0,004414	0,002542	0,0111
35-39	416	375	541	327	0,004155	0,003746	0,005407	0,003274	0,0143
40-44	660	590	849	510	0,006603	0,005901	0,008494	0,005102	0,0198
45-49	1140	1013	1180	762	0,011399	0,010133	0,011801	0,007616	0,0283
50-54	1807	1621	1906	1239	0,018065	0,016214	0,019059	0,01239	0,0403
55-59	2873	2480	2580	1819	0,028728	0,024805	0,025801	0,018186	0,0485
60-64	4575	4131	4173	2998	0,045753	0,041306	0,04173	0,029981	0,0720
65-69	7761	7192	6855	5284	0,077608	0,071919	0,068547	0,052837	0,0974
70-74	12406	12423	10842	9816	0,124057	0,12423	0,108417	0,098163	0,1419
75+	66176	67563	67679	72255	0,661753	0,675625	0,676778	0,722552	0,1570

Table 4. Interpolation And Obtained By Kernel Method Of Male Deaths Numbers And Rates For 2008

AGE GROUP	WEST DEATH NUMBER	EAST DEATH NUMBER	NORTH DEATH NUMBER	SOUTH DEATH NUMBER	WEST DEATH RATE	EAST DEATH RATE	NORTH DEATH RATE	SOUTH DEATH RATE	2008 KERNEL DEATH RATE
	LEVEL2 3,02	LEVEL2 3,02	LEVEL2 3,02	LEVEL2 3,02	LEVEL2 3,02	LEVEL2 3,02	LEVEL2 3,02	LEVEL2 3,02	h=0,025
0	1600	2083	1909	4006	0,016003	0,020829	0,019087	0,040056	0,0383
1-4	215	202	376	581	0,002155	0,002024	0,003755	0,00581	0,0095
5-9	156	119	258	128	0,001558	0,001186	0,002576	0,001282	0,0035
10-14	136	139	211	132	0,001356	0,001392	0,002111	0,001321	0,0031
15-19	275	298	482	187	0,002747	0,002977	0,004824	0,001866	0,0048
20-24	378	423	694	249	0,003778	0,004229	0,006943	0,002492	0,0050
25-29	364	431	700	284	0,00364	0,004312	0,006995	0,00284	0,0059
30-34	408	471	777	405	0,004082	0,004709	0,007772	0,004054	0,0072
35-39	542	597	880	508	0,00542	0,005966	0,008802	0,005078	0,0101
40-44	862	919	1126	809	0,008618	0,009191	0,01126	0,008088	0,0168
45-49	1538	1628	1552	1317	0,015385	0,016279	0,015519	0,013165	0,0246
50-54	2630	3009	2571	2244	0,026296	0,030093	0,025705	0,022444	0,0331
55-59	4497	4957	3432	3530	0,044974	0,049575	0,034322	0,0353	0,0442
60-64	6954	7390	5433	5257	0,069543	0,073902	0,05433	0,052566	0,0603
65-69	10371	10467	8208	7825	0,103708	0,104669	0,082085	0,078253	0,0841
70-74	14398	14523	11592	12218	0,143977	0,145232	0,115924	0,12218	0,1389
75+	54676	52344	59799	60320	0,546757	0,523439	0,597986	0,603202	0,5091

Graphics of the findings from the core forecast and 2003 and 2008 sex range of urban death's standardised age distribution in east, west, north, south model life table's death levels have been shown on Figure 1 (2003 Female), Figure 2 (2003 Male), Figure 3 (2008 Female) and Figure 4 (2008 Male).

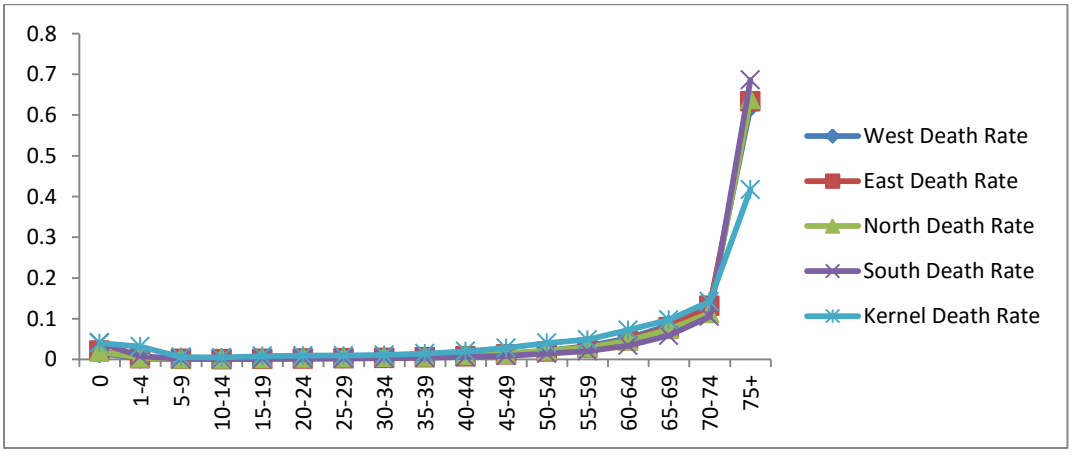

Figure 1. Death Rate of Females for 2003

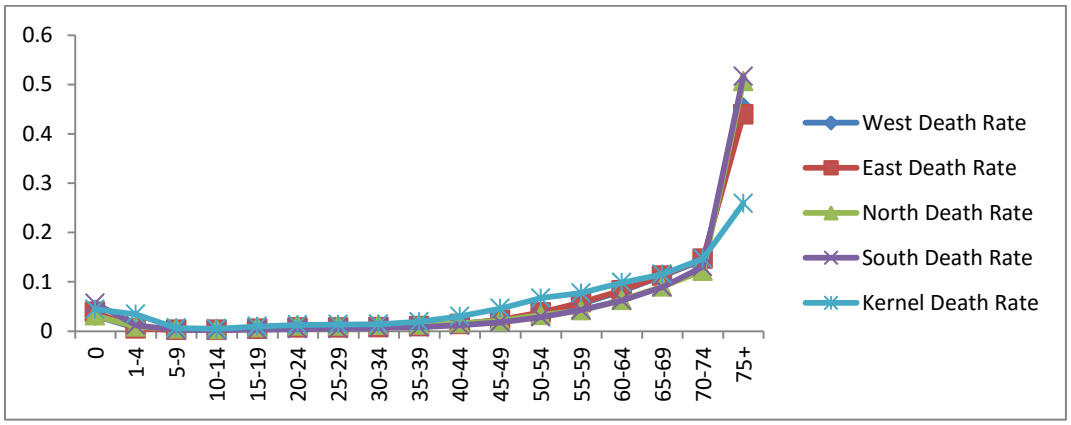

Figure 2. Death Rate of Males for 2003

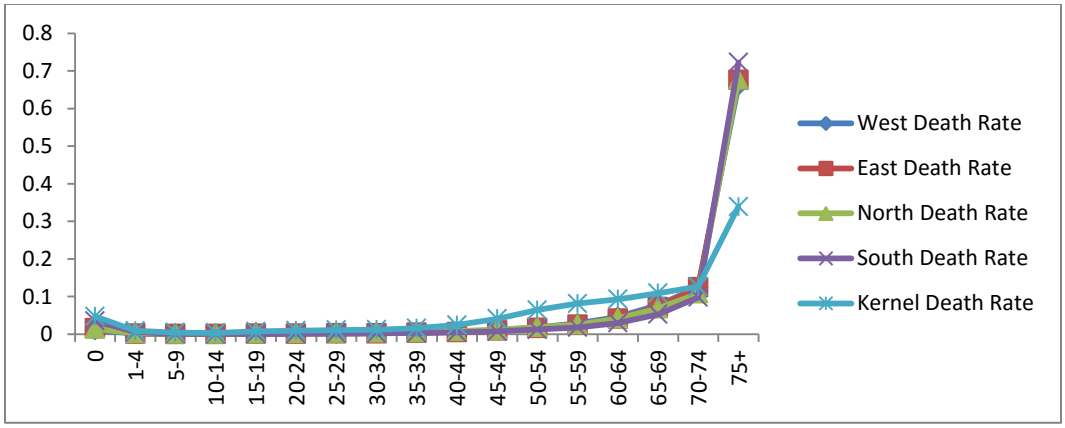

Figure 3. Death Rate of Females for 2008

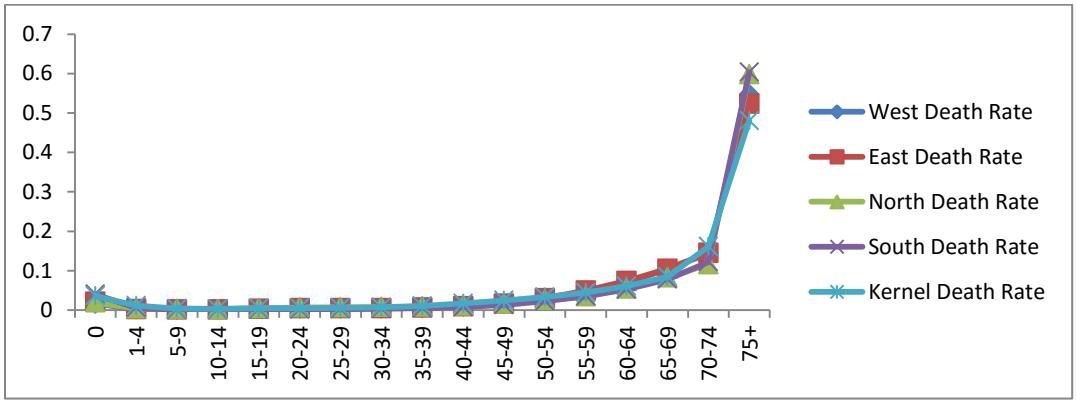

Figure 4. Death Rate of Males for 2008

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. If correlation coefficient (r) is 0,70 or higher, very strong positive relationship; 0,40-0,69, strong positive relationship, 0,30-0,39, moderate positive relationship; 0,20-0,29, weak positive relationship; 0,010-0,19, no or negligible relationship; 0, no relationship. The correlation coefficient for the life table which is obtained from kernel estimation and every standardized age distributed model life tables (west, east, north, south) which are obtained from urban deaths 2003 and 2008 of Turkey by gender. For Turkey **c**orrelation coefficient for 2003 and 2008 are given in the table below according to female and male life cycle (Table 5).

Table 5. Correlation Coefficients for 2003 and 2008

	West-Kernel	East-Kernel	North-Kernel	South-Kernel
Female 2003	0,988	0,986	0,975	0,982
Male 2003	0,955	0,952	0,934	0,946
Female 2008	0,951	0,950	0,917	0,913
Male 2008	0,995	0,994	0,995	0,995

All the correlation coefficients, which are obtained from the analyses between Coale and Demany's model life tables and the age distributions of urban mortalities which are corrected with kernel functions in 2003 and 2008, are found above 0,90. Moreover, the highest correlation coefficient is got from west model for both males and females. These coefficients are 0,988 for females, 0,951 for males in 2003; 0,991 for females, 0,963 for males in 2004; 0,955 for females, 0,995 for males in 2008 in the correlations done with west.

For two discrete probability distributions $P=(p_1, p_2,..., p_k)$ ve $Q=(q_1, q_2,..., q_k)$ their Hellinger distance (H(P,Q)) is defined as

$$H(P,Q)=\frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2} \qquad (2)$$

which is directly related to the Euclidean norm of the difference of the square root vectors, i.e. Hellinger Distance is a distance generally used between two probability distributions. This distance is most suited to detect and measure the anisotropic nature of object parameters.

Hellinger formula that is shown on Equation (2) and that is used in the Science of Statistics to measure the distribution probability closeness to each other was used and the closest distribution has been obtained. Hellinger's information distance for 2003 and 2008 are given in the table below according to female and male life cycle (Table 6).

Table 6. Hellinger Distance

	West-Kernel	East-Kernel	North-Kernel	South-Kernel
Female 2003	0,180	0,190	0,170	0,215
Male 2003	0,293	0,305	0,293	0,330
Female 2008	0,276	0,289	0,277	0,323
Male 2008	0,055	0,061	0,058	0,056

In 2003 according to the female, west-kernel figüre shows 0,180; east-kernel figüre shows 0,190; north-kernel figüre shows 0,170; south-kernel figüre shows 0,215. In 2003 according to the male west-kernel figüre shows 0,293; east-kernel figüre shows 0,305; north-kernel figüre shows 0,293; south-kernel figüre shows 0,330. In 2008 according to the female west-kernel figüre shows 0,276; east-kernel figüre shows 0,289; north-kernel figüre shows 0,277; south-kernel figüre shows 0,323. In 2008 according to the male west-kernel figüre shows 0,055; east-kernel figüre shows 0,061; north-kernel figüre shows 0,058; south-kernel figüre shows 0,056. The lowest Hellinger information distance values are found in the paired comparisons in Table 6 at West-Kernel consistency for males and females for the years 2003 and 2008.

III. Conclusion

Demographs who make researches in Turkey on life tables and mortality use indirect estimation technique. In a study done by Cula and Hosgor it was indicated that adjusted death data can be used with core forecast method while creating life table original to Turkey (Cula and Hosgör, 2010).

In this study, model life table in which urban deaths are converging in Turkey are constructed by using kernel estimation method. The proximities of the age distributions which are obtained from Coale and Demeny's model life tables of different mortality levels and corrected age distributions which are obtained from urban deaths of males and females in 2003 and 2008.

All the correlation coefficients, which are obtained from the analyses between Coale and Demany's model life tables and the age distributions of urban mortalities which are corrected with Kernel functions in 2003 and 2008, are found above 0,90. Moreover, the highest correlation coefficient is got from west model for both males and females. These coefficients are 0,988 for females, 0,951 for males in 2003; 0,991 for females, 0,963 for males in 2004; 0,955 for females, 0,995 for males in 2008 in the correlations done with west.

In the study, Hellinger information distances are used in order to measure the information distance between two probability distributions. The most closest distributions of which the information distances are calculated are given in Table 6. The lowest Hellinger information distance values are found in the paired comparisons in Table 6 at west-kernel consistency for males and females for the years 2003 and 2008.

For future studies, this study can be extended as producing life tables for Turkey by gender and comparing the age distributions of these life tables with the ones of Coale and Demeny's model life tables.

References

Büyüköztürk, Ş. (2010). Sosyal bilimler için veri analizi el kitabı [Data analysis handbook for social sciences] (12nd ed.). Ankara, Turkey: Pegem Yayıncılık.

Coale, A.J., Demeny, P. (1983). Regional Model Life Tables And Stable Populations, Academic Press, New-York.

Cula, G. S., Toktamış, Ö. (2000). Estimation Of Multivariate Probability Density Function With Kernel Functions, Journal Of The Turkish Statistical Association. Vol 3, 29-39.

Cula, S., Hoşgör, Ş. (2010). Application Of Kernel Estimation Method For Correction Age Distribution Errors In Census, Hacettepe Üniversitesi Nüfus Etütleri Nüfusbilim Dergisi Sayı: 28-29.

Hancıoğlu, A. (1996). Unpublished Phd Thesis. Institute Of Population Studies. Hacettepe University. Ankara.

Hoşgör, Ş. (1997). Estimation Of Post-Childhood Life Tables Of Provinces And Regions İn Turkey, By Using Age And Sex Distributions And Intercensal Growth Rates, (1985-1990), Ph.D. Thesis, Institute Of Population Studies Hacettepe University, Ankara.

Hoşgör, Ş. (2005). Türkiye'de Devlet İstatistik Enstitüsünce Derlenen Ölüm Verilerindeki Eksik Kapsamın Boyut Ve Nitelik Tahmini, Turkish Journal Of Population Studies, 27, 17-34.

Hoşgör, Ş. Tansel, A. (2011). Demography And Management Towards 2050, Tüsiad

Kanazawa, Y., (1993). Hellinger distance and Akaike's information criterion for the histogram, Statistics and Probability Letters. Vol. 17, Issue 4, 293-298.

McGraw KO, Wong SP.(1996). Forming inferences about some intraclass correlation coefficients. Psychological Methods;l(1):30-46.

Scot, D.W., Terrell, G.R. (1987). Biased And Unbiased Cross-Validation İn Density Estimation, Journal Of The American Statistical Association, Vol.82, No. 400, 1131-1146.

Simonoff, J. S. (1996). Smoothing Methods İn Statistics. Verlag, New York.

Shryock, H.S., Siegel, J.S. (1988). The Methods And Materials Of Demography. Longman-Press, New-York.

Toros,.A. (1996). Estimation Of Life Expectency Of Turkey, Journal Of Population Studies. Ankara, Turkey.

TUİK (2007). 2003-2004 Death Statistics.

TUİK (2010). 2007-2008 Death Statistics.

U.N, Manual Iv. (1986). New-York.

U.N, Manual X, (1988). New-York.

Wand, M.P., Jones, M.C. (1995). Kernel Smoothing. Chapman & Hall, London.