APPLICATION OF KERNEL ESTIMATION METHOD FOR CORRECTION OF AGE DISTRIBUTION ERRORS IN CENSUS

Serpil CULA*
Şeref HOŞGÖR**

Age is continuous variable. Most of correction techniques assume that it is discrete variable. In this case, to correct age distribution censuses by which techniques have used it, continuous variable is very important. This study aims to discuss if the application of kernel estimation method to age distribution of data acquired in 2000 General Census can be utilized to make correction of misreported age declaration.

INTRODUCTION

Age declarations acquired by censuses of population can be grouped into four categories as errors caused by the participant's missing age information, errors caused by inferential estimation of ages, errors during interviews, and errors during data processing stages. When literatures of Demography considered, it is observed that the methods such as Ariaga, Carrier-Farrag, Karup.King-Newton, United Nations, and Strong moving averages are applied to correct age distribution structure but they only result in minor corrections in the relevant age distribution structure. Furthermore, when the level of faulty age declarations is high, these methods cannot give any satisfactory results. In such cases, Strong moving averages method is generally appeared to be the most applied method. Thus, this study aims to discuss if it is possible to correct misreported age declarations by utilizing Kernel Estimation Method to data acquired in 2000 General Census.

METHOD

One of the nonparametric estimation methods of Regression function is kernel estimation method (Wand and Jones, 1995). There are various methods for kernel estimation of Regression function and in this study Nadaraya-Watson kernel estimation method was used. It is assumed that independent x variable is random variable in Nadaraya-Watson kernel estimation method. This estimator is used in such conditions when both dependent and independent variables are random. Nadaraya-Watson estimator can be given as the following equation (Simonoff, 1996)

$$\hat{m}_{NW}(x) = \frac{\sum_{i=1}^{n} Y_i W(\frac{x - X_i}{h})}{\sum_{i=1}^{n} W(\frac{x - X_i}{h})}$$

where W is a kernel function satisfying $\int W(x)dx = 1$, 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

* Adress correspondence to Serpil Cula (Asst.Prof.), Baskent University, Faculty of Commercial Sciences, Eskişehir Yolu 20.km., 06530 Ankara/TURKEY. TEL:+90(312)2341010/1755 e-mail:scula@baskent.edu.tr
** Adress correspondence to Seref HOŞGÖR (Asst.Prof.), Baskent University, Faculty of Commercial Sciences, Eskişehir Yolu 20.km., 06530 Ankara/TURKEY. TEL:+90(312)2341010/17o2 e-mail:shosgor@baskent.edu.tr

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). However, no method, being commonly accepted, is available yet in spite of most of the studies conducted so far. Some suggested methods are not welcomed very much because of their dependence of unknown density *f* but considered significant to achieve initial values for some methods (Park and Marron, 1990). Each developed method has advantages and disadvantages compared to others (Cula, 1998).

When taking a mass with group number K $(i=1,2,...,K)$ and *i.* observation number for group n_i, total observation number n, and relative frequency for each group, $\overline{p}_i = n_i/n$, where in i/K co-spatial design $\overline{p}_i$'s are reply values, the relative frequency estimators for regression function at each group could be achieved as follows similar to Nadaraya Watson kernel estimation (Simonoff, 1996)

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

In Eq.(2), *W* is kernel function and *h* is bandwidth.

APPLICATION

Accurate age data in census, the ratios of each age group in n total population can be converted to continuous probability distribution that the sum equals 1. This function of distribution shows continuity. It defines the probability function that has wide distribution and the area under the curve Equals 1.

Firstly, the computer programs have been written to obtain the values of kernel estimates. The program has been coded by Delphi 3.0. The least squares cross-validation method has been used to obtain the bandwidth h.

In this study, considering the results of 2000 general census for various age groups, the values of kernel were found.

Standard normal kernel function for *W* kernel function was applied. Besides, least squares cross-validation method for finding *h* bandwidth related to data were applied as well. It is found that separately proper *h* bandwidth for the population data for male and female. *h* bandwidth for female population is 0.048, *h* bandwidth for male population is 0.049. The kernel estimation related to the data achieved by using Eq. (2) putting the estimator place achieved these bandwidths. The real estimation values are calculated by multiplying the achieved estimation values and populations for each age group. The difference between total age distributions achieved by kernel estimations and observed total age distribution equaled by multiplying correction coefficient achieved by the formula is given at Eq. (3). The formula of used correction coefficient *(CC)* is as follows that

$$CC = \frac{OP}{EP}$$

where *OP*=Observed total population, *EP*=Estimated total population.

The graphics of new odd age structures achieved by distributions of kernel estimations are shown in Figure 1 and Figure 2 separately for female and male.

On the other hand, the aim of convergence of initial estimations of both genders to the total population with the *h* window widths has changed. Besides, the estimations achieved by the *h* window width that gave the most approximate results have been discussed by evaluating the other age correction methods.

Figure 1. Kernel estimates and relative frequency of cell probability for male population

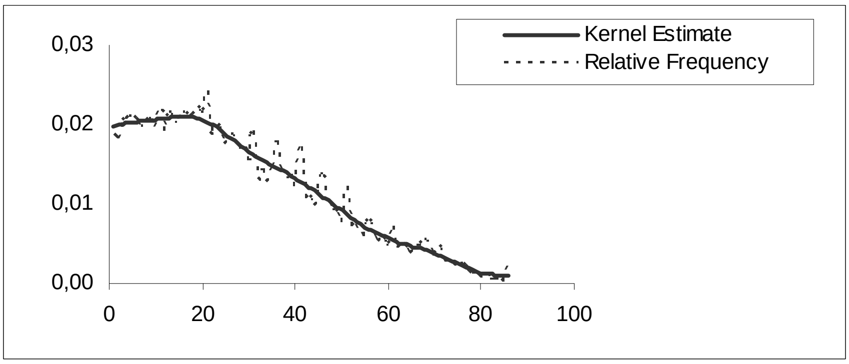
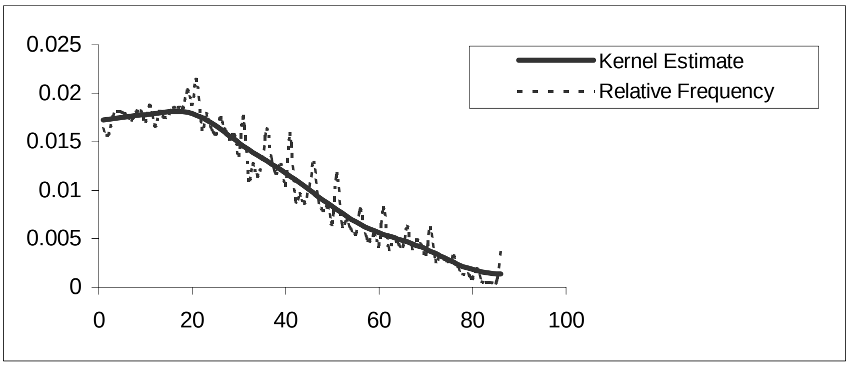


Figure 2. Kernel estimates and relative frequency of cell probability for female population



It is seen from the Figures 1 and 2 that the correction in accurate ages using *W* kernel functions does not depend upon any assumption, 2000 census does not deviate from odd age distribution, and it presents this distribution as equable distribution.

Table 1 shows age distributions groups of 2000 general census by using other correction methods and age distributions groups corrected *W* kernel functions in gender separation together. United Nation's age-gender correction index has applied all corrected age distributions (see Table 1). United Nation's age-gender correction index has also applied the achieved estimations for both genders by using values calculated least square cross-validation method from kernel estimations and age distributions by Eq. (3), the distributions achieved by changing *h* window width to convergent to total observed population. Especially the two estimations that have changed *h* window width and ensuring convergence to total observed population for male h=0.0206 and for female h=0.0315 window width has 13.7 high coefficient according to United Nation's age-gender correction index. This situation changed the age structure so it is decided to use estimated distribution with ideal *h* window width.

The error index coefficients achieved by applying United Nation's age-gender correction index to 2000 census age as well as gender distribution derived by various methods are; for Carrier-Farrag 11.3, for Ariaga 10.6, for K.King- Newton 10.6, for United Nations 10.8, for strong active mean 6.5, for the suggested achieved value in this study (Please see Appendices).

Table 1. Declared and corrected population according to age and gender (2000)

Age Group	Declared Population	Corrected Population					
		Carrier Farrag	K-King Newton	Arriaga	United Nations	Strong moving averages	Cula-Hoşgör
				Men			
0 – 4	3,397,931			3,343,419		3,387,958	3,430,612
5 – 9	3,487,019			3,541,531		3,496,992	3,503,270
10 – 14	3,571,961	3,665,036	3,662,230	3,688,035	3,600,753	3,570,761	3,564,964
15 – 19	3,692,566	3,599,491	3,602,297	3,576,492	3,653,802	3,468,209	3,571,795
20 – 24	3,427,966	3,351,566	3,343,788	3,355,010	3,427,170	3,304,782	3,360,030
25 – 29	2,977,516	3,053,916	3,061,694	3,050,472	2,972,074	3,041,862	2,987,358
30 – 34	2,553,301	2,667,439	2,667,002	2,663,135	2,609,295	2,719,526	2,650,537
35 – 39	2,454,476	2,340,338	2,340,775	2,344,642	2,400,390	2,395,767	2,378,152
40 – 44	2,084,292	2,074,445	2,062,478	2,066,854	2,099,760	2,050,077	2,067,293
45 – 49	1,711,381	1,721,228	1,733,195	1,728,819	1,712,964	1,738,344	1,713,813
50 – 54	1,356,887	1,309,027	1,320,249	1,305,490	1,345,749	1,419,812	1,372,370
55 – 59	1,016,625	1,064,485	1,053,263	1,068,022	1,034,106	1,161,650	1,061,222
60 – 64	864,613	945,565	929,949	933,563	876,148	929,325	876,163
65 – 69	795,171	714,219	729,835	726,221	763,202	706,639	724,029
70 – 74	518,058			504,423		490,796	514,737
75 – 79	254,536			268,171		281,798	300,400
80 +	182,436						269,991
Unknown	12,536						
Total	34,346,735	26,506,755	26,506,755	34,164,299	26,495,414	34,164,299	34,346,735
				Women			
0 – 4	3,189,166			3,142,478		3,174,816	3,222,136
5 – 9	3,271,931			3,318,619		3,286,281	3,276,504
10 – 14	3,309,071	3,432,924	3,431,537	3,452,611	3,362,648	3,366,060	3,333,921
15 – 19	3,519,398	3,395,545	3,396,932	3,375,858	3,456,033	3,287,415	3,346,218
20 – 24	3,264,490	3,223,518	3,215,181	3,229,356	3,289,652	3,161,419	3,193,028
25 – 29	2,919,772	2,960,744	2,969,081	2,954,906	2,885,439	2,923,996	2,891,284
30 – 34	2,458,082	2,590,216	2,588,575	2,586,297	2,538,494	2,625,280	2,578,013
35 – 39	2,401,587	2,269,453	2,271,094	2,273,372	2,325,835	2,319,621	2,296,420
40 – 44	1,985,869	1,982,123	1,975,694	1,976,221	2,017,485	1,983,562	1,992,232
45 – 49	1,658,549	1,662,295	1,668,724	1,668,197	1,658,154	1,697,641	1,667,033
50 – 54	1,361,399	1,306,325	1,316,215	1,302,614	1,341,690	1,412,507	1,358,972
55 – 59	1,042,505	1,097,579	1,087,689	1,101,290	1,076,400	1,185,016	1,116,310
60 – 64	965,302	1,009,738	997,211	1,002,425	950,644	987,854	950,526
65 – 69	850,912	806,476	819,003	813,789	851,531	786,566	798,983
70 – 74	654,985			604,297		587,522	603,287
75 – 79	323,260			373,948		390,723	401,742
80 +	280,914						430,586
Unknown	10,845						
Total	33,457,192	25,736,936	25,736,936	33,176,278	25,754,003	33,176,278	33,457,192

DISCUSSION AND CONCLUSION

When Table 1 is examined, for 0-4 age group estimated Arriaga and strong active mean method estimated less than declared population in this age group. Whereas in demographic literatures lack of ages, especially 0-4 age group and 5 included for wrong age declarations. Therefore, 0-4 age group is less than expected value. Our correction method with *W* kernel functions estimates 0-4 age group more than declared.

On the other hand, the mod age group for both genders occurred 15-19 age group in declared 2000 census, all correction methods except for United Nations age correction method resulted that mod age group has changed 10-14 age group to mod age group. Estimation conducted by *W* kernel function resulted proper estimation of mod age group achieved by census.

Another important manner in here, being age group transitions between 0-14 age both declared ages and Arriaga and strong active mean method, which are increasing amount of and leaped transitions between age groups up to mod group. On the other hand, in the corrections done by *W* kernel function, it is estimated decreasing amount of smooth transitions. Therefore, it makes the trial method more acceptable than the other methods.

The biggest difference from the active mean method was considered through using age population and 10 years period of age group. In other words, taking *h* bandwidth constant in active mean method was noticed. On the other hand, in W kernel function estimation method, *h* bandwidth is changeable and the ratio of age population in total census was applied. In other words, probabilities were used. Accordingly, it gives opportunity to decide ideal *h* bandwidth in the ratio of convergence to total population at the same time. As a matter of fact, in this study *h* window width has changed for elimination the difference estimated total population and observed total population but at the end it is decided that the age distribution has been deteriorated by the result of applied United Nations age-gender correction index, which is higher, and this method has not been used. But it is impossible to say that the method shall increase error index continuously in different populations in distribution of age and gender.

The used five methods above in age correction brought out discussion of not to make sufficient correction in 10- year- age group, the used methods utilize various convergence probabilities for this age group. The changing *h* window width using ability of kernel functions makes it possible to reach as many as approximate results in initial and final ages.

The estimated method suggested by Cula-Hoşgör with the comparison of five different age correction methods, as the result of applied United Nations age-sex accuracy index in age distributions grouped by genders. The error index coefficient 8.6 in Cula-Hoşgör method is lower than calculated error index coefficients in all methods except for strong active mean. Applying changing of age distribution mod group, it is assumed that the age variable is discontinuous variable in Strong active mean method as well as this unchanging mod group. Moreover, it is assumed that the age variable is continuous variable, which is an important advantage of the validity of method.

As a result, age is continuous variable that has unit time and divided into so small pieces. Many correction methods experienced and used by now have considered the age variable as discontinuous variable due to its structure and applying methods. But in this study, continuous age variable declaration errors were divided by *n* number of kernel functions as its *h* window width and the areas under the curve for each function and correction of *W* kernel function estimation method, which is evaluating the weighted probabilities of other areas in total made. It seems that it is a proper method when continuous variable is corrected by the continuous function.

REFERENCES

Canpolat, Ş. (2003), Türkiye'de Gerçekleştirilen Nüfus Sayımlarında Yaş Bildirim Kalitesinin Analizi. Uzmanlık Tezi, D.İ.E., Ankara.

Cula, S.C. (1998), Çok Değişkenli Olasılık Yoğunluk Fonksiyonunun Çekirdek Fonksiyonlarıyla Kestirimi. Doktora Tezi.

Cula, G. S. and Toktamış, Ö. (2000), Estimation of Multivariate Probability Density Function With Kernel Functions. **Journal of the Turkish Statistical Association,** 3, 29-39.

Horova, I. And Zelinka, J. (2007), Contribution to the bandwidth choice for kernel density estimates, **Computational Statistics,** 22, 31-47.

Hoşgör, Ş. (1997), Estimation of Post-Childhood Life Tables of Provinces and regions in Turkey, by Using Age and Sex Distributions and Intercensal Growth Rates, (1985-1990). Ph.D. Thesis, Institute of Population Studies Hacettepe University, Ankara.

Jones, M. C. and Marron, J.S. (1996), A brief survey of bandwidth selection for density estimation. **Journal of the American Statistical Association,** 91, Issue 433, 401-407p.

Müller, H.G., Wang, J.L., Carey, J.R., Caswell-Chen, E.P., Chen, C., Papadopoulos, C. Yao, F. (2004), Demographic window to aging in the wild: constructing life tables and estimating survival functions from marked individuals of unknown age, **Aging Cell.** June, 3(3), 125-131.

Park, B. U., Marron, J.S., (1990), Comparison of data-driven bandwidth selectors. **J. Am. Stat. Assoc,** 85, 66-72.

Rosenblatt, M. (1956), Remarks on some Nonparametric Estimates a Density Function. **Annals Math. Statististics,** 27, 832-837

Shryock, H.S. and Siegel, J.S. (1988), **Methods and Materials of Demography.** Longman-Press.

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

Silverman, B. W. (1986), **Density Estimation for Statistics and Data Analysis.** Chapman and Hall, London.

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

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

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

APPENDICES

Appendix A: UNITED NATIONS AGE SEX ACCURACY INDEX; CARRIER-FARRAG

UNITED NATIONS AGE SEX ACCURACY INDEX; CARRİER-FARRAG											
	POPULATION		ANALYSIS OF SEX RATIO			ANALYSIS OF AGE RATIO					
	MALE	FEMALE	RATIO	SUC DIF	ABSUCDF	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
						RATIO	RATIO	DF 100	DF 100	ABDF100	ABDF100
10-14	3,665,036	3,432,924	106.76	X	X	X	X	X	X	X	X
15-19	3,599,491	3,395,545	106.01	0.8	0.8	102.6	102.0	2.6	2.0	2.6	2.0
20-24	3,351,566	3,223,518	103.97	2.0	2.0	100.7	101.4	0.7	1.4	0.7	1.4
25-29	3,053,916	2,960,744	103.15	0.8	0.8	101.5	101.9	1.5	1.9	1.5	1.9
30-34	2,667,439	2,590,216	102.98	0.2	0.2	98.9	99.0	-1.1	-1.0	1.1	1.0
35-39	2,340,338	2,269,453	103.12	-0.1	0.1	98.7	99.3	-1.3	-0.7	1.3	0.7
40-44	2,074,445	1,982,123	104.66	-1.5	1.5	102.2	100.8	2.2	0.8	2.2	0.8
45-49	1,721,228	1,662,295	103.55	1.1	1.1	101.7	101.1	1.7	1.1	1.7	1.1
50-54	1,309,027	1,306,325	100.21	3.3	3.3	94.0	94.7	-6.0	-5.3	6.0	5.3
55-59	1,064,485	1,097,579	96.98	3.2	3.2	94.4	94.8	-5.6	-5.2	5.6	5.2
60-64	945,565	1,009,738	93.64	3.3	3.3	106.3	106.1	6.3	6.1	6.3	6.1
65-69	714,219	806,476	88.56	5.1	5.1	X	X	X	X	X	X
TOTAL					21.4					29.0	25.5
MEAN					1.9					2.9	2.6
United nations age-sex accuracy index =					**11.3**						

Appendix B: UNITED NATIONS AGE SEX ACCURACY INDEX; ARIAGA

UNITED NATIONS AGE SEX ACCURACY INDEX; CARRİER-FARRAG											
	POPULATION		ANALYSIS OF SEX RATIO			ANALYSIS OF AGE RATIO					
	MALE	FEMALE	RATIO	SUC DIF	ABSUCDF	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
						RATIO	RATIO	DF 100	DF 100	ABDF100	ABDF100
10-14	3,688,035	3,452,611	106.82	X	X	X	X	X	X	X	X
15-19	3,576,492	3,375,858	105.94	0.9	0.9	101.6	101.0	1.6	1.0	1.6	1.0
20-24	3,355,010	3,229,356	103.89	2.1	2.1	101.3	102.0	1.3	2.0	1.3	2.0
25-29	3,050,472	2,954,906	103.23	0.7	0.7	101.4	101.6	1.4	1.6	1.4	1.6
30-34	2,663,135	2,586,297	102.97	0.3	0.3	98.7	98.9	-1.3	-1.1	1.3	1.1
35-39	2,344,642	2,273,372	103.13	-0.2	0.2	99.1	99.7	-0.9	-0.3	0.9	0.3
40-44	2,066,854	1,976,221	104.59	-1.5	1.5	101.5	100.3	1.5	0.3	1.5	0.3
45-49	1,728,819	1,,668,197	103.63	1.0	1.0	102.5	101.8	2.5	1.8	2.5	1.8
50-54	1,305,490	1,302,614	100.22	3.4	3.4	93.4	94.1	-6.6	-5.9	6.6	5.9
55-59	1,068,022	1,101,290	96.98	3.2	3.2	95.4	95.6	-4.6	-4.4	4.6	4.4
60-64	933,563	1,002,425	93.13	3.8	3.8	104.1	104.7	4.1	4.7	4.1	4.7
65-69	726,221	813,789	89.24	3.9	3.9	X	X	X	X	X	X
TOTAL					21.0					25.8	23.1
MEAN					1.9					2.6	2.3
United nations age-sex accuracy index =					**10.6**						

Appendix C: UNITED NATIONS AGE SEX ACCURACY INDEX; K.KING-NEWTON

UNITED NATIONS AGE SEX ACCURACY INDEX; CARRİER-FARRAG											
	POPULATION		ANALYSIS OF SEX RATIO			ANALYSIS OF AGE RATIO					
	MALE	FEMALE	RATIO	SUC DIF	ABSUCDF	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
						RATIO	RATIO	DF 100	DF 100	ABDF100	ABDF100
10-14	3,662,230	3,431,537	106.72	X	X	X	X	X	X	X	X
15-19	3,602,297	3,396,932	106.05	0.7	0.7	102.8	102.2	2.8	2.2	2.8	2.2
20-24	3,343,788	3,215,181	104.00	2.0	2.0	100.4	101.0	0.4	1.0	0.4	1.0
25-29	3,061,694	2,969,081	103.12	0.9	0.9	101.9	102.3	1.9	2.3	1.9	2.3
30-34	2,667,002	2,588,575	103.03	0.1	0.1	98.7	98.8	-1.3	-1.2	1.3	1.2
35-39	2,340,775	2,271,094	103.07	0.0	0.0	99.0	99.5	-1.0	-0.5	1.0	0.5
40-44	2,062,478	1,975,694	104.39	-1.3	1.3	101.3	100.3	1.3	0.3	1.3	0.3
45-49	1,733,195	1,668,724	103.86	0.5	0.5	102.5	101.4	2.5	1.4	2.5	1.4
50-54	1,320,249	1,316,215	100.31	3.6	3.6	94.8	95.5	-5.2	-4.5	5.2	4.5
55-59	1,053,263	1,087,689	96.83	3.5	3.5	93.6	94.0	-6.4	-6.0	6.4	6.0
60-64	929,949	997,211	93.25	3.6	3.6	104.3	104.6	4.3	4.6	4.3	4.6
65-69	729,835	819,003	89.11	4.1	4.1	X	X	X	X	X	X
TOTAL					20.3					27.1	24.0
MEAN					1.8					2.7	2.4
United nations age-sex accuracy index =					**10.6**						

Appendix D: UNITED NATIONS AGE SEX ACCURACY INDEX; UNITED NATIONS

UNITED NATIONS AGE SEX ACCURACY INDEX; CARRİER-FARRAG											
	POPULATION		ANALYSIS OF SEX RATIO			ANALYSIS OF AGE RATIO					
	MALE	FEMALE	RATIO	SUC DIF	ABSUCDF	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
						RATIO	RATIO	DF 100	DF 100	ABDF100	ABDF100
10-14	3,600,753	3,362,648	107.08	X	X	X	X	X	X	X	X
15-19	3,653,802	3,456,033	105.72	1.4	1.4	104.0	103.9	4.0	3.9	4.0	3.9
20-24	3,427,170	3,289,652	104.18	1.5	1.5	103.4	103.8	3.4	3.8	3.4	3.8
25-29	2,972,074	2,885,439	103.00	1.2	1.2	98.5	99.0	-1.5	-1.0	1.5	1.0
30-34	2,609,295	2,538,494	102.79	0.2	0.2	97.1	97.4	-2.9	-2.6	2.9	2.6
35-39	2,400,390	2,325,835	103.21	-0.4	0.4	101.9	102.1	1.9	2.1	1.9	2.1
40-44	2,099,760	2,017,485	104.08	-0.9	0.9	102.1	101.3	2.1	1.3	2.1	1.3
45-49	1,712,964	1,658,154	103.31	0.8	0.8	99.4	98.7	-0.6	-1.3	0.6	1.3
50-54	1,345,749	1,341,690	100.30	3.0	3.0	98.0	98.1	-2.0	-1.9	2.0	1.9
55-59	1,034,106	1,076,400	96.07	4.2	4.2	93.1	93.9	-6.9	-6.1	6.9	6.1
60-64	876,148	950,644	92.16	3.9	3.9	97.5	98.6	-2.5	-1.4	2.5	1.4
65-69	763,202	851,531	89.63	2.5	2.5	X	X	X	X	X	X
TOTAL					20.0					27.8	25.4
MEAN					1.8					2.8	2.5
United nations age-sex accuracy index =					**10.8**						

Appendix E: UNITED NATIONS AGE SEX ACCURACY INDEX; MOVING AVERAGE

UNITED NATIONS AGE SEX ACCURACY INDEX; CARRİER-FARRAG											
	POPULATION		ANALYSIS OF SEX RATIO			ANALYSIS OF AGE RATIO					
	MALE	FEMALE	RATIO	SUC DIF	ABSUCDF	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
						RATIO	RATIO	DF 100	DF 100	ABDF100	ABDF100
10-14	3,570,761	3,366,060	106.08	X	X	X	X	X	X	X	X
15-19	3,468,209	3,287,415	105.50	0.6	0.6	100.9	100.7	0.9	0.7	0.9	0.7
20-24	3,304,782	3,161,419	104.53	1.0	1.0	101.5	101.8	1.5	1.8	1.5	1.8
25-29	3,041,862	2,923,996	104.03	0.5	0.5	101.0	101.1	1.0	1.1	1.0	1.1
30-34	2,719,526	2,625,280	103.59	0.4	0.4	100.0	100.1	0.0	0.1	0.0	0.1
35-39	2,395,767	2,319,621	103.28	0.3	0.3	100.5	100.7	0.5	0.7	0.5	0.7
40-44	2,050,077	1,983,562	103.35	-0.1	1.0	99.2	98.8	-0.8	-1.2	0.8	1.2
45-49	1,738,344	1,697,641	102.40	1.0	1.0	100.2	100.0	0.2	0.0	0.2	0.0
50-54	1,419,812	1,412,507	100.52	1.9	1.9	97.9	98.0	-2.1	-2.0	2.1	2.0
55-59	1,161,650	1,185,016	98.03	2.5	2.5	98.9	98.7	-1.1	-1.3	1.1	1.3
60-64	929,325	987,854	94.08	4.0	4.0	99.5	100.2	-0.5	0.2	0.5	0.2
65-69	706,639	786,566	89.84	4.2	4.2	X	X	X	X	X	X
TOTAL					17.4					8.6	9.1
MEAN					1.6					0.9	0.9
United nations age-sex accuracy index =					**6.5**						

Appendix F: UNITED NATIONS AGE SEX ACCURACY INDEX; CULA-HOŞGÖR

UNITED NATIONS AGE SEX ACCURACY INDEX; CARRİER-FARRAG											
	POPULATION		ANALYSIS OF SEX RATIO			ANALYSIS OF AGE RATIO					
	MALE	FEMALE	RATIO	SUC DIF	ABSUCDF	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
						RATIO	RATIO	DF 100	DF 100	ABDF100	ABDF100
10-14	3,572,825	3,347,389	106.73	X	X	X	X	X	X	X	X
15-19	3,582,906	3,394,471	105.55	1.2	1.2	103.2	103.1	3.2	3.1	3.2	3.1
20-24	3,369,568	3,234,960	104.16	1.4	1.4	102.5	102.8	2.5	2.8	2.5	2.8
25-29	2,992,789	2,899,211	103.23	0.9	0.9	99.4	99.8	-0.6	-0.2	0.6	0.2
30-34	2,654,807	2,575,866	103.06	0.2	0.2	98.8	98.8	-1.2	-1.2	1.2	1.2
35-39	2,383,954	2,313,072	103.06	0.0	0.0	100.9	101.0	0.9	1.0	0.9	1.0
40-44	2,072,097	2,002,515	103.47	-0.4	0.4	101.0	100.5	1.0	0.5	1.0	0.5
45-49	1,717,482	1,670,770	102.80	0.7	0.7	100.3	99.6	0.3	-0.4	0.3	0.4
50-54	1,354,052	1,350,807	100.24	2.6	2.6	97.4	97.4	-2.6	-2.6	2.6	2.6
55-59	1,061,697	1,103,178	96.24	4.0	4.0	95.2	95.8	-4.8	-4.2	4.8	4.2
60-64	877,432	951,288	92.24	4.0	4.0	98.1	99.3	-1.9	-0.7	1.9	0.7
65-69	726,658	813,425	89.33	2.9	2.9	X	X	X	X	X	X
TOTAL					18.3					19.0	16.7
MEAN					1.7					1.9	1.7
United nations age-sex accuracy index =					**8.6**						

ÖZET

NÜFUS SAYIMLARINDAKİ YAŞ DAĞILIM HATALARININ DÜZELTİLMESİ İÇİN KERNEL TAHMİN YÖNTEMİNİN UYGULAMASI

Bu çalışmada, 2000 genel nüfus sayımındaki yaş verilerine, çekirdek kestirim yöntemi, uygulayıp hatalı yaş beyanlarının düzeltilmesinde bu yöntemin kullanılıp kullanılamayacağı tartışılmıştır. Bu amaçla çekirdek kestirim yöntemi ile düzeltilen yaş verilerine, literatürde belirtilen ve sıkça kullanılan Ariaga, Carrier-Farag, K.King-Newton, Birleşmiş Milletler, Hareketli ortalamalar, gibi yaş düzeltme teknikleri uygulanmış, 2000 nüfus sayımı verisi ile Birleşmiş Milletler Yaş -Cinsiyet hata indeksi kullanılarak karşılaştırılmıştır. Karşılaştırma sonucunda, uygulanan tekniğin sıralamada hata indeksi yönünden en iyi ikinci sonucu verdiği görülmüştür. Bu nedenle çekirdek kestirim yöntemi, nüfus sayımlarındaki hatalı yaş dağılımlarını düzeltmek için uygun ve kabul edilebilir bir yöntem olduğu kanaati ile tartışmaya açılmış ve kullanıcılara yeni bir yöntem olarak sunulmuştur.