APPLICATION OF SYNTHETIC ESTIMATION TO SELECTED DEMOGRAPHIC AND HEALTH INDICATORS AS A TRADITIONAL SMALL AREA ESTIMATION FOR PROVINCES OF TURKEY

Ahmet Sinan TÜRKYILMAZ *

Demographic and Health Surveys provide accurate and detailed demographic information and some basic health information. However, this information is limited to nation totals, urban/rural levels and at most to regions due to the nature of sample surveys. Recently, the use of survey data in developing reliable small area statistics, possibly in relation to the census and survey data, has received attention. In this study "Synthetic estimation techniques" were used as an example of traditional small area methods. The aim of this study is to describe and to present the results of synthetic estimates for eighty provinces of Turkey for 1998. The 1990 population census and 1998 Turkish Demographic and Health Survey were used as data sources. When the results are considered, it can be seen that provincial estimates did not change dramatically from the direct regional estimates.

INTRODUCTION

As a developing country, Turkey does not have a complete and reliable registration system on statistical, demographic and health data. Turkey has been taking censuses since the beginning of the establishment of the Republic. As well known, censuses provide demographic information as detailed as settlement level, but this information is very limited and has some deficiencies. It is not possible to get useful information from the registration system even for nation as a whole. The existing registration systems have some coverage, reliability and accuracy problems. Sample surveys provide accurate and detailed demographic information and some basic health information. However, this information is limited to nation totals, urban/rural levels and at most to geographical regions due to the nature of sample surveys.

In recent years there has been a substantial growth in the demand for statistical data relating to various geographical subdivisions within the country. The geographical subdivisions of interest sometimes comprise relatively large units, such as provinces, or states, and they sometimes comprise smaller units, such as towns, rural communities, local government districts, or health service areas. Statistics for geographical subdivisions, commonly referred to as small area statistics, are of great interest in many countries throughout the world (Kalton, et. al, 1993).

The use of small area statistics originated several centuries ago. Brackstone (1987) mentions the existence of such statistics in 11th century England and 17th century in Canada. Fortunately, with the high-speed computers, fast processing of large data sets made feasible the provision of timely data for the small areas. As well, several powerful statistical methods with theoretical foundations have emerged for the analysis of local data. Indirect Small Area estimation techniques can be classified into two main groups (Rao, 2000, Marker 1999): Traditional Techniques and Model Based Techniques. Synthetic Estimates are considered as one the Traditional Techniques and this technique requires relatively available data from the surveys and censuses.

* Dr., Instructor, Hacettepe University, Institute of Population Studies

There are very limited studies done in Turkey, reviewing or using small area estimation techniques. There are some master thesis including a methodological review and a case study. Özcan (1991), Sevinç (1998) and Deliloğlu (2001) performed studies as master thesis on small area estimation techniques. Türkyılmaz (2003) studied small area estimation techniques for Turkey in his Ph. D. dissertation. The 1990 population census and 1998 Turkish Demographic and Health Survey were used as data sources. Four small area estimation techniques were used in this dissertation. A technique was developed as an adaptation of "multiple imputation technique". The main idea is to impute the estimates of districts where the census variables assumed the fully observed variables and survey variables assumed to be variables with missing in an aggregated district level data set. Systematically selected demographic and health indicators were estimated by combining the observed and imputed district level estimates for provincial level. It is possible to calculate standard errors with multiple imputation technique. A third series of estimates, called composite estimates, were also calculated. This estimate is a combination of direct and estimates of multiple imputations whose variances minimized. Synthetic estimation techniques were used as an example of traditional methods. The aim of this paper is to describe the technique and to present the results of synthetic estimates for eighty provinces of Turkey for 1998.

LITERATURE

Gonzalez (1973) describes the synthetic estimation as follows: "An unbiased estimate is obtained from sample survey for a large area; when this estimate is used to derive estimates for sub-areas under the assumption that the small areas have the same characteristics as the large area, we identify these estimates as synthetic estimates."

Synthetic estimation is different from the symptomatic or sample regression methods which are traditional techniques, it begins with national or regional subgroup estimates, for example age-sex groups then derives small area estimates by taking the appropriate weighted average of national subgroup estimates, where the weights reflect the age-sex composition of the small area. The accuracy of this method depends on how similar each population subgroup's national or regional average is to its small area average, and on the accuracy of the weights.

Marker (1999) gives the synthetic estimator of the average of characteristic Y for small area *i* as:

$$\hat{\bar{Y}}_{i..} = \frac{\left(\sum_j N_{ij}\bar{y}_{.j.}\right)}{N_{i.}}$$

where N_{ij} denotes the size of the total population in small area i, $i=1,2,\ldots,I$, and subgroup j, $j=1,\ldots,J$, $N_{i.}$ is the total population in small area i, and $\bar{y}_{.j.}$ is the sample average of Y for subgroup j, across all small areas.

National Center for Health Statistics (1968) first used synthetic estimates to calculate state estimates of long and short-term physical disabilities from the National Health Interview Survey data. This method is generally used for small area estimation, mainly because of its simplicity, applicability to general sample designs and potential of increased accuracy in estimation borrowing information from similar small areas.

Marker (1999) notes Gonzalez and Waksberg (1973) derive a general variance estimate for the synthetic estimators. They suggest using the average mean square error (MSE) for all domain estimates, and using it, show synthetic estimates to be preferable to regional averages for estimating the errors in vacancy rates in the 1970 U.S. Census. Their MSE estimate gives only an estimate of average error over all domains; there is no way to drive an estimate of the accuracy of any single small area estimate from this procedure.

Purcell and Kish (1979) observed that synthetic estimates are biased estimates for two reasons. First, there will often exist departures from the underlying assumption of homogeneity of rates. Second, the weights are usually based on past data, and the structure of the population may have changed during this time.

Marker (1999) lists papers using these small area estimation techniques, including those by Aliaga and Le (1991), DiGaetano et al. (1980), Gonzalez and Hoza (1978), Ghangurde and Singh (1978), Levy and French (1977), Namekata, Levy and O'Rourke (1975), Schaible (1980) and Shpiece (1981). DiGaetano et al. and Nemhaka et al. both examined using synthetic estimation for the National Health Interview Survey. The Aliaga and Le, DiGaetano et al., Gonzalez and Hoza, and Levy French papers compared synthetic estimation with other small area techniques.

METHODOLOGY AND DATA SOURCE

Synthetic estimation methods are based on an assumption that the small area is similar to another larger area. The method assumes that the estimated rate is roughly constant within larger areas. What is different between the larger and smaller areas are the proportion of individuals in the smaller area who have selected characteristics related to the estimate. Synthetic estimation uses sample data to estimate, at some higher level of aggregation, the variable of interest for different subclasses of population. Then it scales these estimates in proportion to the subclass incidence within the small domains of interest.

Synthetic estimation uses auxiliary data, for example on the distribution of respondents by age or educational attainment, which are correlated with the study variable. The population is first divided into sub-groups according to the auxiliary variable. The information about the relationship between the auxiliary variable and the study variable can be obtained from the survey sample at the larger area level. Then, an estimate is calculated for each sub-group in the small area by weighting the sub-group estimate for the larger area according to the number of cases in the small area. The sum of these sub-group estimates across all categories of the auxiliary variable produces the synthetic estimates for the sub area.

The number of auxiliary variables may increase the consistency of the estimates, but it may also decrease consistency. Since the auxiliary variable is calculated from an external data source, generally a census, the alternatives are limited for auxiliary variables. Although, there are examples of using a single auxiliary variable, the number of auxiliary variables can be increased, and cross relations between more than one variable can be studied.

Aliaga and Le (1991) and Aliaga and Muhuri (1994) studied contraceptive use by using demographic and health survey (DHS) data. They used age as an auxiliary variable in the study. Most of the studies done in developing countries generally consider one auxiliary variable. Gonzalez and et al. (1996) used race, age, and live birth order as auxiliary variables to give synthetic estimates of some health indicators by using vital registration system information and National Natality National Fetal Mortality Survey data in the US.

Following tables are developed to illustrate the calculation procedure and formulas of synthetic estimation of a statistics proportion denoted by "p". Here provinces of a region are the "small areas" and the region that they belong to is the larger area. These tables divide the auxiliary variable into five or three sub-groups. Table 2. stands for a province within a region whereas Table 1. stands for all provinces in a region.

Table 1. Synthetic Estimation with one auxiliary variable

	Auxiliary Variable						
Province	**Sub-group 1**	**Sub-group 2**	**Sub-group 3**	**Sub-group 4**	**Sub-group 5**	**Total**	**Synthetic Estimate**
Province 1	N_{11}	N_{12}	N_{13}	N_{14}	N_{15}	$N_{1.}$	$p_{1.}$
Province 2	N_{21}	N_{22}	N_{23}	N_{24}	N_{25}	$N_{2.}$	$p_{2.}$
Province 3	N_{31}	N_{32}	N_{33}	N_{34}	N_{35}	$N_{3.}$	$p_{3.}$
Province 4	N_{41}	N_{42}	N_{43}	N_{44}	N_{45}	$N_{4.}$	$p_{4.}$
Province 5	N_{51}	N_{52}	N_{53}	N_{54}	N_{55}	$N_{5.}$	$p_{5.}$
.							.
.							.
Region1	$P_{.1}$	$P_{.2}$	$P_{.3}$	$P_{.4}$	$P_{.5}$	N	

The synthetic estimation procedure with one auxiliary variable can be expressed as follows:

$$p_{g.} = \sum_{j} \frac{N_{gj}}{N_{g.}} P_{.j}$$

Here $p_{g.}$ is the synthetic estimate of the statistic "p" in the g^{th} province, $P_{.j}$ is the observed estimate for the j^{th} subgroup of auxiliary variable in the region, generally obtained from a sample survey, N_{gj} is the number of persons in the j^{th} subgroup of the g^{th} province, $N_{g.}$ is the number of persons in the g^{th} province, and $N_{gi} / N_{g.}$ is the adjustment weight.

Table 2. Synthetic Estimation with two auxiliary variables

Province 1				
		Auxiliary Variable 1		
Auxiliary Variable 2	**Sub-group 1**	**Sub-group 2**	**Sub-group 3**	**Total**
Sub-group 1	N_{11}	N_{12}	N_{13}	$N_{1.}$
Sub-group 2	N_{21}	N_{22}	N_{23}	$N_{2.}$
Sub-group 3	N_{31}	N_{32}	N_{33}	$N_{3.}$
Total	$N_{.1}$	$N_{.2}$	$N_{.3}$	N_{1}

Region 1				
		Auxiliary Variable 1		
Auxiliary Variable 2	**Sub-group 1**	**Sub-group 2**	**Sub-group 3**	**Region 1**
Sub-group 1	P_{11}	P_{12}	P_{13}	$P_{1.}$
Sub-group 2	P_{21}	P_{22}	P_{23}	$P_{2.}$
Sub-group 3	P_{31}	P_{32}	P_{33}	$P_{3.}$
Region 1	$P_{.1}$	$P_{.2}$	$P_{.3}$	P_{1}

The synthetic estimation procedure with two auxiliary variables can be expressed as:

$$p_g = \sum_{ij} \frac{N_{gj}}{N_g} P_{ij}$$

Here p_g is the synthetic estimate in the g^{th} province, P_{ij} is the observed estimate for the ij^{th} subgroup of auxiliary variables in the region, N_{gj} is the number of persons in the ij^{th} subgroup of the g^{th} province who are denominator for the statistic "p", N_g is the total number of persons who are denominator for the statistic "p" in the g^{th} province, and N_{gj} / N_g is the adjustment weight.

Auxiliary Variable Construction: Previous examples of synthetic estimation, which work with the demographic and health survey, generally used only the age distribution of the respondents. In this study the estimates from the larger areas, which are the five regions, have reliable estimates from the 1998 TDHS. The survey provides the regional estimates for sub-groups of the variables.

The source for the adjustment weights is the 1990 Population Census. More than one auxiliary variable is considered in order to improve the reliability of the estimates. The selection of auxiliary variables depends on the census variables rather than the variables in the survey. Although there are many variables in the survey, the census provides limited common variables to be used as auxiliary ones. Further an auxiliary variable should be a strong relation to the variable of interest. Second must also be available for all sub-groups of both the survey and the census.

Age, of course was considered first, and then other variables of the census were considered. It was decided that educational level is important factor which may determine the main relation between the study variables and settlements. Except for two variables based on household information, all other variables are based on information coming from women aged 15-49. Naturally, there is an obvious relation between the characteristic of a woman and the household in which she lives.

After selecting age and educational level as two auxiliary variables, the question was how to determine the sub-groups of these auxiliary variables. Aliaga and Le (1991) used a five sub-group approach for the women aged 15-49, and Aliaga and Muhuri (1994) used a three sub-group approach for the women aged 15-49. The underlying reason for categorizing the women in aged 15-49 to a three groups is a logistic regression study, which showed the differences between women, aged 15-24, 25-34 and 35-49. It appeared more rational to use fewer categories for forming sub-groups. Another reason was the number of cases in the survey for each category within the region. Increasing the number of sub-group leads to empty cells in cross tables for regions, or cells based on very few observations.

A study was performed to see the outcomes of the different categorization approaches. In order to see these different approaches, a simulation was done using census data. The main idea of this simulation is to produce census based direct estimates and comparing them with census based synthetic estimates under different categorization schemes. It is possible to produce some indicators, like children ever born, from census data alone. Normally regional estimates are calculated from sample data and adjustment weights are calculated from census data. In the simulation, the five regional estimates were created using the census data, and, instead of regional base sample information, regional census information was used. Adjustment weights were also calculated from census data.

Age and education variables were considered and three different “education” auxiliary variable and two different “age” auxiliary variable categories were created for women aged 15-49. The last number in the names of the variable denotes the number of categories in each variable. Synthetic estimates were computed one or two auxiliary variables. The interest variables were “children ever born”, and the “proportion of ever married women who did not worked last week”. Different approaches were calculated for:

- Age3
- Age7
- Edu2
- Edu3
- Edu7
- Edu2*Age3
- Edu3*Age3
- Edu7*Age3

Provincial based direct estimates were also calculated for the indicators. Then all synthetic estimates results were compared with provincial direct estimates. Absolute differences were calculated. It was noticed that two auxiliary variable categories produce better results, and one with fewer categories used the Edu3*Age3 categories.

The 14 subregional base estimates for synthetic estimates were also considered. When the estimates were calculated for 14 subregions it was seen that some cells in the age*education subgroups were empty, especially for subregions of West region. When the all figures were calculated for five regions it was also observed that provincial base estimates do not differ from direct regional estimates, as expected. So, five regions were used as larger areas.

Two main data sources were used in this study: The 1990 Turkish Population Census and 1998 Turkey Demographic and Health Survey (TDHS). In addition, the 1997 Turkish Population Count also provided some general information to identify settlements at the time of 1998 TDHS.

The 1990 Population Census, which was the thirteenth population census since the foundation of the Turkish Republic, was carried out by the State Institute of Statistics (SIS) on 21st of October 1990. The unit for the addresses is “dwelling” and the unit for enumeration is “individual person”. The aim of the census was to determine precisely population size and social and economical characteristics of the domestic and alien population living in provinces, districts, sub-districts, and villages within the country’s boundaries on the census day (SIS, 1993).

The 1998 Turkey Demographic and Health Survey (TDHS-98) is the seventh in a series of national-level population and health surveys conducted during the last thirty years. The primary objective of the TDHS-98 is to provide data on fertility and mortality, family planning, maternal and child health, and reproductive health. The survey obtained detailed information on these issues from a sample of women in the reproductive ages (15-49) and from the husbands of currently married eligible women (HUIPS and MEASURE/DHS+, 1999). A disproportionately, multistage, stratified cluster sampling approach was used in the selection of the TDHS-98 sample (Türkyılmaz and Aliaga, 1999). The sample was designed in this fashion because of the need to provide estimates for a variety of characteristics for various domains. These domains, which are frequently employed in the tabulation of major indicators from the survey, are: Turkey as a whole, Urban and rural areas (each as a separate domain), each of the major five regions of the country. These five regions are a combination of 14 subregions those were created for stratification purposes. A total of the 8,059 households were successfully interviewed. In the interviewed households, Interviews were successfully completed with 8,576 of the women aged 15-49. In half of the selected households, 1,971 husbands were successfully interviewed. Sampling errors of the survey were presented in appendices of the survey report (Aliaga and Türkyılmaz, 1999).

The indicator selection was made from a very long list of possible indicators. Combining different sets of indicators from different sources, one list was the list of the indicators or variables produced from 1998 TDHS. A second list was the UNICEF's "Indicators for Global Monitoring of Child Rights", UNICEF 1998), a third list was the list of indicators of "Quality of Life Indicators on Women and Children" (Hancıoğlu, Koç and Dayıoğlu, 2000). A fourth list is Ünalan's "Population and Development Indicators" list (Ünalan, 2002). We also paid a special attention to the indicators in the cover of the 1998 TDHS report (HUIPS and MEAUSRE+/DHS, 1999), those named under UNICEF's "World Summit for Children Indicators" which also is a part of "Indicators for Monitoring Progress at End-Decade", and sub set of indicators than the UN's "Human Development Index" (UN, 2002). A subset on related topics was taken from the "Indicators for Global Monitoring of Child Rights". "Quality of Life Indicators on Women and Children" is a list of 25 indicators be produced from 1998 TDHS. Ünalan's list is a combination of different sets of indicators on "Population and Development" produced by ICPD, UNFPA-IPRHP, UNFPA-CPA, MNSDS, BSSA, UNDAF/CCA, OECD/DAC, UNCSD, MDG, FoC Tier 1. A combined list of these different sources made a list of approximately 200 indicators. After a selection procedure, following indicators were selected as "study" indicators or "key" indicators as proportion type statisitics:

1. Households using safe water
2. Households using sanitary toilet
3. Infant mortality rate (0-4 years)
4. Antenatal Care
5. Births delivered at health facilities
6. Mothers received medical care at birth
7. Had diarrhea in the last 2 weeks
8. Underweight Children (Weight-for-age, below -2 SD)
9. Children under 5 years of age without birth registration
10. Consanguinity
11. Currently using a modern contraceptive method
12. Proportion of deceased Children

THE RESULTS

The Results of Synthetic Estimation estimates are presented in this section. As it was described, synthetic estimates are based on the regional level estimates of the 12 "key" variables. The factor that was changing the regional estimates is the structural differences of auxiliary variables. In other words, the auxiliary variables of this study are the three-category age and three-category education variable of the women, if the age*education distribution of the province is very similar to the age*education distribution of the region that it belongs to, than the provincial estimate is not different than regional estimates.

Table 3. Direct Estimates for "Key Variables" by five-regions of 1998 TDHS

	Region					
Indicator	**1-West**	**2-South**	**3-Centre**	**4-North**	**5-East**	**Turkey**
1. Households using safe water	0.77	0.78	0.78	0.56	0.67	0.75
2. Households using sanitary toilet	0.94	0.86	0.84	0.85	0.78	0.88
3. Infant mortality rate (0-4 years)	0.032	0.030	0.041	0.044	0.061	0.042
4. Antenatal Care	0.86	0.72	0.74	0.67	0.38	0.68
5. Births delivered at health facilities	0.87	0.70	0.84	0.84	0.45	0.73
6. Mothers received medical care at birth	0.92	0.86	0.90	0.90	0.52	0.81
7. Had diarrhea in the last 2 weeks	0.22	0.29	0.31	0.26	0.40	0.30
8. Underweight Children (Weight-for-age, below -2 SD)	0.04	0.09	0.05	0.05	0.17	0.08
9. Children under 5 years of age without birth registration	0.12	0.18	0.20	0.13	0.42	0.22
10. Consanguinity	0.16	0.30	0.27	0.24	0.38	0.25
11. Currently using a modern contraceptive method	0.41	0.35	0.43	0.35	0.27	0.38
12. Proportion of deceased Children	0.048	0.050	0.074	0.084	0.094	0.068

When the results are considered, it can be seen that provincial estimates did not change dramatically from the regional estimates. The regional direct estimates are given in Table 3. As it was mentioned, it is not possible to produce the standard error estimates of synthetic estimates so it is hard to conclude the reliability of the estimates. Here, results are given for 12 "key" variables in Table 4, in which the results are sorted for the provinces within five regions of TDHS. It can easily be seen from the Table 4. The results are very similar within region and when we compared the estimates with regional direct estimates, significant differences are not observed.

When the provincial results of synthetic estimates are examined, it is seen that the estimates for "Households using safe water" indicator has estimates of minimum value of 51 percent for Sinop and maximum of 82 percent for Ankara, the smallest estimates are all for North Region provinces, followed by East region provinces. It can be observed that estimates generally differ within region in a range of 5 –7 percent. When the "Households using sanitary toilet" indicator is examined, the minimum estimates are for provinces of East region with a smallest value of 72 percent. Maximum estimates are calculated for provinces of West region. These results are nothing but, approximately ± 5 percent of the regional estimates. The same conclusion can be made for the Infant Mortality Rate.

Similarly, the synthetic estimates of "Antenatal Care", "Births delivered at health facilities", "Mothers received medical care at birth", "Had diarrhea in the last 2 weeks", "Underweight

Children” and “Children under 5 years of age without birth registration” indicators have estimated values very parallel to the regional estimates. The differences between regional estimates are reflected in the differences between provinces, the provinces from same regions give closer estimates as expected.

The results for the indicators “Consanguinity”, “Currently using a modern contraceptive method”, and “Proportion of deceased Children” are the same, when the provincial synthetic estimates are examined. The minimum estimate for “Consanguinity” indicator for provinces of West region is 0.14 and maximum is 0.16 where the regional direct estimate is 0.16. Larger differences are seen within the East region provinces, for instance Şırnak has an estimate of 47 percent where the regional estimate is 38 percent. Very similar estimates are observed to give same comments for “Currently using a modern contraceptive method”, and “Proportion of deceased Children” indicators.

CONCLUSION

The estimates of synthetic estimation was done under the assumption of “provinces are similar to regions that they belong to and what differs provinces is the cross proportion of age and educational groups of provinces in a region”. It is also assumed that age and educational level of women are deterministic for all indicators. The selection and then definition of auxiliary variables, which are three category age and educational level, was done under a simulation study. This study showed that, although there were no dramatic differences, “three-category with two auxiliary variables selection” would give better results.

The results of synthetic estimates are not that much different than regional estimates as expected. It is not possible to calculate standard errors for this method so it is not possible to measure reliability. However it can be concluded that, if there is no estimation available or only unreliable estimates are available for a small area, a province in our case, synthetic estimates can be used. Under this assumption, if a subregional direct estimate is reliable and available, this value can also be used if there are concerns about the relation between regional and provincial similarities and province is believed to be more similar to subregion.

The use of subregional estimates were also considered to be used instead of regions in the synthetic estimates but it was seen that some of the subregional estimates are very unreliable especially for the cross groups of age and education, or there are empty cells for such groups. Then it was only decided to make suggestions that invite the use of reliable subregional direct estimates for provinces if no other estimates are available.

Table 4. Synthetic Estimates for All "Key variables", sorted within region

	Province	Households Using Safe water	Households Using Sanitary Toilet	Infant Mortality Rate	Antenatal Care	Births Delivered at health Facilities	Medical Care at birth
1	AYDIN	0.75	0.93	0.036	0.86	0.86	0.91
1	BALIKESİR	0.76	0.93	0.035	0.87	0.86	0.92
1	BURSA	0.77	0.94	0.034	0.87	0.87	0.92
1	ÇANAKKALE	0.75	0.93	0.035	0.87	0.86	0.92
1	DENİZLİ	0.76	0.93	0.036	0.86	0.86	0.91
1	EDİRNE	0.76	0.93	0.034	0.87	0.87	0.92
1	İSTANBUL	0.80	0.95	0.033	0.87	0.87	0.92
1	İZMİR	0.78	0.95	0.033	0.87	0.87	0.92
1	KIRKLARELİ	0.77	0.94	0.032	0.89	0.88	0.93
1	KOCAELİ	0.78	0.95	0.035	0.86	0.86	0.91
1	MANİSA	0.75	0.93	0.038	0.84	0.84	0.90
1	SAKARYA	0.76	0.94	0.037	0.85	0.85	0.91
1	TEKİRDAĞ	0.77	0.94	0.032	0.88	0.88	0.93
1	YALOVA	0.79	0.95	0.031	0.89	0.88	0.93
2	ADANA	0.79	0.86	0.024	0.69	0.65	0.84
2	ANTALYA	0.78	0.86	0.022	0.74	0.69	0.87
2	BURDUR	0.77	0.86	0.020	0.75	0.70	0.88
2	GAZİANTEP	0.77	0.85	0.030	0.63	0.59	0.79
2	HATAY	0.77	0.85	0.027	0.67	0.62	0.81
2	ISPARTA	0.78	0.86	0.020	0.76	0.71	0.89
2	İÇEL	0.79	0.87	0.022	0.72	0.68	0.86
2	MUĞLA	0.78	0.86	0.020	0.76	0.72	0.89
2	KİLİS	0.75	0.83	0.028	0.65	0.59	0.79
2	OSMANİYE	0.77	0.85	0.027	0.67	0.62	0.81
3	AFYON	0.77	0.82	0.058	0.68	0.81	0.86
3	AMASYA	0.76	0.81	0.060	0.67	0.81	0.85
3	ANKARA	0.82	0.86	0.056	0.75	0.85	0.89
3	BİLECİK	0.77	0.82	0.056	0.72	0.84	0.89
3	BOLU	0.76	0.81	0.060	0.67	0.81	0.85
3	ÇANKIRI	0.75	0.80	0.059	0.69	0.82	0.86
3	ÇORUM	0.75	0.79	0.061	0.64	0.79	0.83
3	ESKİŞEHİR	0.79	0.85	0.057	0.73	0.85	0.89
3	KAYSERİ	0.78	0.82	0.059	0.68	0.81	0.85
3	KIRŞEHİR	0.77	0.82	0.057	0.69	0.82	0.86
3	KONYA	0.78	0.82	0.059	0.68	0.81	0.86
3	KÜTAHYA	0.76	0.81	0.058	0.68	0.81	0.86
3	NEVŞEHİR	0.77	0.82	0.057	0.68	0.81	0.86
3	NİĞDE	0.76	0.81	0.060	0.65	0.80	0.84
3	TOKAT	0.75	0.79	0.062	0.64	0.79	0.83
3	UŞAK	0.77	0.81	0.059	0.68	0.81	0.86

Table 4. Synthetic Estimates for All "Key variables", sorted within region (continued)

	Province	Households Using Safe water	Households Using Sanitary Toilet	Infant Mortality Rate	Antenatal Care	Births Delivered at health Facilities	Medical Care at birth
3	YOZGAT	0.75	0.79	0.061	0.63	0.78	0.83
3	AKSARAY	0.76	0.80	0.061	0.63	0.78	0.82
3	KARAMAN	0.76	0.81	0.057	0.69	0.82	0.87
3	KIRIKKALE	0.79	0.83	0.057	0.69	0.82	0.86
4	ARTVİN	0.55	0.82	0.029	0.60	0.82	0.89
4	GİRESUN	0.53	0.80	0.035	0.60	0.81	0.89
4	KASTAMONU	0.52	0.80	0.032	0.59	0.81	0.89
4	ORDU	0.54	0.80	0.036	0.59	0.81	0.88
4	RİZE	0.57	0.82	0.035	0.59	0.81	0.88
4	SAMSUN	0.55	0.80	0.035	0.60	0.81	0.89
4	SİNOP	0.51	0.79	0.031	0.59	0.81	0.89
4	TRABZON	0.57	0.82	0.035	0.61	0.82	0.89
4	ZONGULDAK	0.58	0.82	0.034	0.61	0.82	0.89
4	BARTIN	0.53	0.79	0.035	0.59	0.81	0.88
4	KARABÜK	0.57	0.82	0.028	0.60	0.82	0.90
5	ADIYAMAN	0.68	0.78	0.057	0.38	0.44	0.52
5	AĞRI	0.66	0.76	0.057	0.34	0.40	0.48
5	BİNGÖL	0.66	0.77	0.057	0.37	0.43	0.51
5	BİTLİS	0.67	0.77	0.056	0.35	0.41	0.49
5	DİYARBAKIR	0.67	0.78	0.056	0.34	0.41	0.48
5	ELAZIĞ	0.71	0.81	0.051	0.43	0.49	0.57
5	ERZİNCAN	0.71	0.82	0.047	0.49	0.55	0.63
5	ERZURUM	0.70	0.80	0.051	0.43	0.49	0.57
5	GÜMÜŞHANE	0.70	0.80	0.049	0.48	0.53	0.62
5	HAKKARİ	0.66	0.77	0.058	0.31	0.38	0.45
5	KARS	0.71	0.81	0.050	0.44	0.51	0.58
5	MALATYA	0.71	0.81	0.048	0.48	0.54	0.61
5	K.MARAŞ	0.69	0.80	0.054	0.42	0.49	0.56
5	MARDİN	0.66	0.76	0.058	0.33	0.40	0.47
5	MUŞ	0.66	0.77	0.058	0.33	0.40	0.47
5	SİİRT	0.66	0.76	0.057	0.33	0.40	0.47
5	SİVAS	0.70	0.80	0.051	0.46	0.52	0.60
5	TUNCELİ	0.69	0.79	0.050	0.46	0.51	0.59
5	ŞANLIURFA	0.67	0.77	0.057	0.34	0.40	0.48
5	VAN	0.66	0.77	0.056	0.34	0.40	0.48
5	BAYBURT	0.69	0.80	0.054	0.44	0.51	0.59
5	BATMAN	0.67	0.77	0.058	0.34	0.40	0.48
5	ŞIRNAK	0.62	0.72	0.059	0.29	0.36	0.43
5	ARDAHAN	0.69	0.79	0.051	0.45	0.52	0.60
5	IĞDIR	0.67	0.77	0.054	0.38	0.44	0.52

Table 4. Synthetic Estimates for All “Key variables”, sorted within region (continued)

	Province	Had diarrhea in the last 2 weeks	Underweight Children	Children under 5 without birth registration	Consan-guinity	Using a modern contraceptive	Proportion of deceased Children
1	AYDIN	0.25	0.03	0.12	0.15	0.47	0.036
1	BALIKESİR	0.25	0.03	0.11	0.15	0.48	0.037
1	BURSA	0.24	0.03	0.11	0.15	0.47	0.037
1	ÇANAKKALE	0.25	0.03	0.10	0.15	0.48	0.035
1	DENİZLİ	0.25	0.03	0.11	0.16	0.47	0.037
1	EDİRNE	0.24	0.03	0.11	0.15	0.47	0.035
1	İSTANBUL	0.23	0.03	0.10	0.15	0.46	0.038
1	İZMİR	0.23	0.03	0.10	0.14	0.47	0.037
1	KIRKLARELİ	0.24	0.02	0.10	0.15	0.47	0.036
1	KOCAELİ	0.24	0.03	0.11	0.15	0.46	0.037
1	MANİSA	0.26	0.03	0.13	0.15	0.47	0.037
1	SAKARYA	0.25	0.03	0.12	0.15	0.47	0.038
1	TEKİRDAĞ	0.24	0.03	0.10	0.15	0.47	0.037
1	YALOVA	0.23	0.03	0.09	0.15	0.47	0.038
2	ADANA	0.29	0.09	0.17	0.26	0.37	0.041
2	ANTALYA	0.27	0.08	0.16	0.25	0.40	0.040
2	BURDUR	0.27	0.08	0.17	0.25	0.41	0.040
2	GAZİANTEP	0.32	0.11	0.19	0.27	0.34	0.040
2	HATAY	0.30	0.10	0.18	0.26	0.36	0.041
2	ISPARTA	0.27	0.08	0.16	0.25	0.41	0.040
2	İÇEL	0.28	0.08	0.17	0.25	0.39	0.040
2	MUĞLA	0.27	0.08	0.17	0.25	0.41	0.039
2	KİLİS	0.31	0.10	0.18	0.26	0.35	0.041
2	OSMANİYE	0.30	0.10	0.18	0.26	0.36	0.041
3	AFYON	0.31	0.07	0.20	0.22	0.47	0.060
3	AMASYA	0.31	0.07	0.20	0.22	0.47	0.059
3	ANKARA	0.27	0.06	0.18	0.22	0.47	0.061
3	BİLECİK	0.28	0.07	0.19	0.22	0.49	0.061
3	BOLU	0.31	0.07	0.20	0.22	0.47	0.058
3	ÇANKIRI	0.30	0.07	0.20	0.22	0.47	0.062
3	ÇORUM	0.33	0.08	0.21	0.22	0.46	0.060
3	ESKİŞEHİR	0.27	0.07	0.19	0.22	0.49	0.061
3	KAYSERİ	0.31	0.07	0.21	0.22	0.46	0.061
3	KIRŞEHİR	0.31	0.07	0.20	0.23	0.46	0.064
3	KONYA	0.30	0.07	0.20	0.22	0.47	0.062
3	KÜTAHYA	0.31	0.07	0.20	0.22	0.47	0.060
3	NEVŞEHİR	0.31	0.07	0.20	0.23	0.46	0.062
3	NİĞDE	0.32	0.08	0.22	0.22	0.46	0.062
3	TOKAT	0.33	0.08	0.22	0.22	0.45	0.059
3	UŞAK	0.31	0.07	0.20	0.22	0.47	0.060

Table 4. Synthetic Estimates for All "Key variables", sorted within region (continued)

	Province	Had diarrhea in the last 2 weeks	Underweight Children	Children under 5 without birth registration	Consan-guinity	Using a modern contraceptive	Proportion of deceased Children
3	YOZGAT	0.34	0.08	0.22	0.23	0.45	0.060
3	AKSARAY	0.35	0.08	0.22	0.23	0.44	0.060
3	KARAMAN	0.30	0.07	0.20	0.22	0.48	0.063
3	KIRIKKALE	0.30	0.07	0.20	0.22	0.47	0.063
4	ARTVİN	0.24	0.04	0.13	0.22	0.39	0.067
4	GİRESUN	0.24	0.04	0.15	0.22	0.37	0.059
4	KASTAMONU	0.24	0.04	0.14	0.22	0.39	0.061
4	ORDU	0.23	0.04	0.15	0.22	0.37	0.059
4	RİZE	0.24	0.04	0.14	0.22	0.38	0.065
4	SAMSUN	0.24	0.04	0.14	0.22	0.38	0.060
4	SİNOP	0.24	0.04	0.14	0.22	0.39	0.062
4	TRABZON	0.24	0.04	0.14	0.22	0.38	0.063
4	ZONGULDAK	0.25	0.04	0.13	0.22	0.38	0.061
4	BARTIN	0.24	0.04	0.14	0.22	0.38	0.058
4	KARABÜK	0.25	0.04	0.12	0.22	0.40	0.064
5	ADIYAMAN	0.36	0.16	0.39	0.34	0.33	0.053
5	AĞRI	0.37	0.18	0.42	0.36	0.30	0.054
5	BİNGÖL	0.37	0.17	0.41	0.35	0.31	0.054
5	BİTLİS	0.37	0.18	0.42	0.36	0.30	0.056
5	DİYARBAKIR	0.37	0.19	0.43	0.36	0.29	0.056
5	ELAZIĞ	0.37	0.14	0.36	0.32	0.35	0.051
5	ERZİNCAN	0.37	0.11	0.32	0.29	0.38	0.048
5	ERZURUM	0.36	0.13	0.36	0.31	0.36	0.049
5	GÜMÜŞHANE	0.37	0.12	0.33	0.30	0.38	0.046
5	HAKKARİ	0.38	0.20	0.46	0.38	0.27	0.058
5	KARS	0.37	0.13	0.35	0.31	0.36	0.048
5	MALATYA	0.37	0.12	0.32	0.30	0.37	0.049
5	K.MARAŞ	0.37	0.14	0.37	0.32	0.35	0.050
5	MARDİN	0.37	0.19	0.43	0.37	0.29	0.055
5	MUŞ	0.37	0.18	0.43	0.36	0.29	0.054
5	SİİRT	0.37	0.19	0.44	0.37	0.29	0.054
5	SİVAS	0.37	0.12	0.34	0.31	0.37	0.048
5	TUNCELİ	0.37	0.12	0.34	0.31	0.37	0.049
5	ŞANLIURFA	0.37	0.18	0.42	0.36	0.30	0.056
5	VAN	0.37	0.19	0.43	0.36	0.29	0.055
5	BAYBURT	0.38	0.12	0.35	0.31	0.36	0.049
5	BATMAN	0.37	0.18	0.43	0.37	0.29	0.056
5	ŞIRNAK	0.37	0.21	0.47	0.38	0.27	0.057
5	ARDAHAN	0.37	0.12	0.34	0.30	0.38	0.046
5	IĞDIR	0.36	0.16	0.38	0.33	0.33	0.050

REFERENCES

Aliaga A. and Muhuri R.K., 1994. **Methods of Estimating Contraceptive Prevalence Rates for Small Areas: Applications in the Dominican Republic and Kenya.** DHS Methodological Report No:3, Maryland, USA.

Aliaga A. and Le T., 1991.**Methodology for Small Area Estimation with DHS Samples,** in Sample Design and Supplementation Session in Demographic and Health Survey World Conference Proceedings, Volume I, 497-512.

Aliaga, A. and Türkyılmaz, A.S., 1999. **App. C: Estimates of Sampling Errors, In Turkish Demographic and Health Survey 1998**, Ankara: Hacettepe University Institute of Population Studies and Measure DHS+ Macro International Inc., 171-181.

Brackstone, G. J., 1987. **Small Area Data: Policy Issues and Technical Challenges**, in Small Area Statistics, New York: John Wiley and Sons.

Deliloğlu R.A.S, 2001. **Small Domain Estimation Techniques for Census and Survey Data**.M.Sc. Thesis, Middle East Technical University, Department of Statistics.

DiGaetano et al., 1980. *Synthetic* **Estimates for Local Areas from Health Interview Survey**. Proceeding of the Section on Survey Research Methods, American Statistical Association, 46-55.

Ghangurde, P.D. and Singh, M.P., 1978. **Use and evaluation of Synthetic Estimates**. Proceedings of the Social Statistics Section, American Statistical Association, 33-36.

Gonzalez, M. and Hoza, C., 1978. **Small Area Estimation with Application to Unemployement and Housing Estimates.** Journal of the American Statistical Association, 73, 7-15.

Gonzalez, M.E., 1973. **Use and evaluation of synthetic estimators**. In Proceedings of the Social Statistics Section. 33-36. Amer. Statist. Assoc., Washington, DC.

Hacettepe University Institute of Population Studies (HUIPS) and MEASURE DHS+ Macro International, 1994. *1998* **Turkey Demographic and Health Survey**. Hacettepe Institute of Population Studies, 1999.

Hancıoğlu, A., Koç I., and Dayıoglu, M., 2000, **The Progress of Regions in Turkey 2000 Towards advancing the Quality of Life Indicators on Women and Children**. Draft Report, UNICEF.

Human Development Report 2001 web page http://hdr.undp.org/reports/global/2001/en/

Kalton, G., Kordos, J. and Platek,R. (Eds), 1993. **Small Area Statistics and Survey Designs Vol.I: Invited Papers; Vol. II: Contributed Papers and Panel Discussion**. Central Statistical Office, Warsaw.

Levy, P.S. and French, D.K., 1977. **Synthetic Estimates of State Health Characteristics Based on the Health Interview Survey**. Vital and Health Statistics, 2:75.

Marker D. A., 1999. **Organization of Small Area Estimators Using a Generalized Linear Regression Framework**, Journal of Official Statistics, 15(1),1-24.

Namekata et al., 1975. **Synthetic Estimates of work Loss Disability for Each State and the District of Columbia**. Public Health Reports, 90(6), 532-538.

National Center for Health Statistics, 1968. **Synthetic State Estimates of Disability**. P.H.S. Publication 1759.U.S. Government Printing Office, Washington DC.

Özcan C., 1991. **Small Domain Estimation Techniques for Official Statistics Based on Sample Survey**.M.Sc. Thesis, Middle East Technical University, Department of Statistics.

Purcell, N.J. and Kish, L., 1979. **Estimation for Small Domains**. Biometrics, 35, 165-384.

Rao, J.N.K, 2000a. **Statistical Methodology for Indirect Estimations in Small Areas**. Seminaro International De Estadisdica En Euskadi, 39, ISBN:84-7749-350-2.

Schaible, W.L., 1980. **A Discussion of "Synthetic Estimates for Local Areas from Health Interview Survey"** by DiGaetano et al. Proceeding of the Section on Survey Research Methods, American Statistical Association, 56.

Sevinç V, 1998. **Küçük Alan Kestirim Teknikleri ve Bolu İli İşsizlik Oranının Kestirimi Üzerine Bir Uygulama**. Master Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı.

Shpiece, M.R., 1981. **The Use of Synthetic Estimation Estimating the Elderly Population in Need: A Study and Comments**. Proceedings of the Social Statistics Section, American Statistical Association, 267-272.

State Institute of Statistics (SIS), 1993. *1990* **Census of Population: Social and Economic Characteristics of Population**. State Institute of Statistics, Printing Division, Ankara.

Türkyılmaz, A.S. and A. Aliaga, 1999**. App. B: Survey Design, In Turkish Demographic and Health Survey 1998**, Ankara: Hacettepe University Institute of Population Studies and Measure DHS+ Macro International Inc., 153-168.

Türkyılmaz, A.S. (2003), Unpublished Ph.D. Dissertation, **Estimation of Selected Demographic and Health Indicators for Provinces of Turkey from Census and Survey Data by Using Small Area Estimation Techniques**, Department of Technical Demography, Hacettepe University Institute of Population Studies, May 2003, Ankara, Turkey.

U.S. Bureau of the Census, 1966. **Methods of Population Estimation: Part I, Illustrative Procedure of the Bureau's Component Method II** . Current Population Reports, Series P-25. No:339. Washington DC.

Ünalan, T., 2002. **Population and Development Indicators: National and International Practices**. TUBITAK and UNFPA, Population and Development Strategies Subprogramme Area, Ankara.

UNICEF, 1998. **Indicators for Global Monitoring of Child Rights**. International Meeting sponsored by UNICEF, Geneva, Switzerland, Summary Report.UNICEF, New York

United Nations, 2002. **Human Development Report 2001**. UN publications.

ÖZET

SENTETİK TAHMİN YÖNTEMİNİN GELENEKSEL BİR KÜÇÜK ALAN KESTİRİMİ OLARAK TÜRKİYE'NİN İLLERİNDE SEÇİLMİŞ NÜFUS VE SAĞLIK GÖSTERGELERİNE UYGULANMASI

Nüfus ve Sağlık araştırmaları detaylı ve geçerli demografik veri ve bazı temel sağlık göstergeleri için bilgi sağlamaktadır. Bununla birlikte, bu bilgi araştırmaların doğası gereği ülke toplamı, kır/kent ve en fazla bölge düzeyindedir. Son zamanlarda, araştırma verisinin sayım verisi ile birlikte küçük alanlar için güvenilir bilgi üretmesi dikkat çekmektedir. Bu araştırmada "Sentetik tahmin tekniği" geleneksel küçük alan tekniklerine bir örnek olarak kullanılmıştır. Bu çalışmanın amacı tekniği açıklamak 1998 yılı için Türkiye'nin seksen ili bazında hesaplanan sonuçları sunmaktır. 1990 genel nüfus sayımı ve 1998 Türkiye Nüfus ve Sağlık araştırması veri kaynağı olarak kullanılmıştır. Sonuçlara bakıldığında, il bazı tahminlerin bölgeler için yapılan doğrudan tahminlerden dramatik olarak farklılaşmadığı görülmüştür.