THE EGYPTIAN POPULATION AND FAMILY PLANNING REVIEW.
ISSR,CAIRO UNIV. Vol.34,No.1,2001

Median Information Using Auxiliary Information

by
Mounira A. Hussein*

In sample surveys, supplementary information is often used for increasing the precision of the estimators of population parameters. The resulting methods of estimation, which make use of the ancillary information, under certain conditions give more efficient estimates of the population parameters than those based on the information of the predicted variable alone (Hussein, 1989). The author had earlier suggested position and stratified median estimators for stratified samples (Hussein, 1998). The author had also suggested a separate and combined ratio estimators (Hussein, 1999). Most of the researches of median estimation deals exclusively with the survey variable of interest alone and do not use the auxiliary variables in the construction of the estimators (Kuk and Mak 1989, Rao and shao 1996). This paper deals with the estimation, under simple random sampling of a finite population median in the presence of more than one auxiliary variables. The two proposed estimator are used to estimate the median water per capita in water short countries.

I. Position Estimator in the presence of more than one auxiliary variables:

The estimator defined here attempts to utilize two auxiliary variables X_{i1} , X_{i2} to construct a p with smaller expected mean square error and consequently smaller variance for estimating the population median.

Consider the three-way classification:

	$X_1 \le M_{X_1}$, $X_2 \le M_{X_2}$	$X_1 \le M_{X_1}$, $X_2 > M_{X_2}$	$X_1 > M_{X_1}$, $X_2 \le M_{X_2}$	$X_1 > M_{X_1}$, $X_2 > M_{X_2}$
$Y \le M_T$	P_{111}	P_{112}	P_{121}	P_{122}
$Y > M_T$	P_{211}	P_{212}	P_{221}	P_{222}
	$P_{.11}$	$P_{.12}$	$P_{.21}$	$P_{.22}$

1. Dr. Mounira Ahmed Hussein is associate professor, college of commerce, Menoufia University.

Let $Y_{(1)} \le Y_{(2)} \le ... \le Y_{(n)}$ be the ordered Y values in S_n. Let $n_{x_1x_2}$ be the number of units in S_n with $X_1 \le M_{X_1}$ and $X_2 \le M_{X_2}$ Let n_{x_1} be the number on units in S_n with $X_1 \le M_{X_1}$ and $X_2 > M_{X_2}$. Let n_{x_2} be the number of units in S_n with $X_1 > M_{X_1}$ and $X_2 \le M_{X_2}$.

So, we can propose a position estimator of P as an extension of the estimation given by Kuk and Mak (1988) as follows:

$$\hat{P}_B = n^{-1}\left\{ n_{x_1x_2}\frac{P_{111}}{P_{.11}} + n_{x_1}\frac{P_{112}}{P_{.12}} + n_{x_2}\frac{P_{121}}{P_{.21}} \right.$$

$$\left. + (n - n_{x_1} - n_{x_2} - n_{x_1x_2})\frac{P_{122}}{P_{.22}} \right\}$$

Thus an estimator of M_Y is alternatively given by $M_{BY} = Q_Y(P_B)$
Where Q_Y is the commutative distribution function.

Since

$$P_{.11} = \frac{n_{x_1x_2}}{n}, \quad P_{.12} = \frac{n_{x_1}}{n}$$

$$P_{.21} = \frac{n_{x_2}}{n}, \quad P_{.22} = \frac{n - n_{x_1} - n_{x_2} - n_{x_1x_2}}{n}$$

$$\therefore V(\hat{P}_B) = \frac{1-f}{n}\{n_{x_1x_2}P_{111}(1-P_{111}) + n_{x_1}P_{112}(1-P_{112})$$

$$+ n_{x_2}P_{121}((1-P_{121}) + (n - n_{x_1} - n_{x_2} - n_{x_1x_2})P_{122}(1-P_{122})\}$$

where $f = \frac{n}{N}$

Considering the first auxiliary variable alone

	$X_1 \le M_{X_1}$	$X_1 > M_{X_1}$
$Y \le M_Y$	P_{11}	P_{12}
$Y > M_Y$	P_{21}	P_{22}
	$P_{.1}$	$P_{.2}$

$$\hat{P}_{B_1} = n^{-1}\left\{ n_X \frac{P_{11}}{P_{.1}} + (n - n_{X_1})\frac{P_{12}}{P_{.2}} \right\}$$

$$V(\hat{P}_{B_1}) = \frac{1-f}{n}\{n_X\ P_{11}(1-P_{11}) + (n - n_X)P_{12}(1-P_{12})\}$$

II. Comparison of the Position Estimator in the Presence of Two Auxiliary Variables with that of One Auxiliary Variable.

Since $V\left(\hat{P}_{B_1}\right)=\frac{1-f}{n}\left\{n_x \; P_{11}\left(1-P_{11}\right)+\left(n-n_x\right)P_{12}\left(1-P_{12}\right)\right\}$

and $n_x = n_{x_1x_2}+n_{x_1}$

$$n-n_x = n_{x_2}+\left(n-n_{x_1}-n_{x_2}-n_{x_1x_2}\right)$$

$$\therefore V\left(\hat{P}_{B_1}\right)=\frac{1-f}{n}\{n_{x_1x_2}+n_{x_1})\left(P_{111}+P_{112}\right)\left(1-P_{111}-P_{112}\right)$$

$$+\left(n_{x_2}+\left(n-n_{x_1}-n_{x_2}-n_{x_1x_2}\right)\left(P_{121}+P_{122}\right)\left(1-P_{121}-P_{122}\right)\right\}$$

$$=\frac{1-f}{n}\{n_{x_1x_2}P_{111}(1-P_{111})+n_{x_1}\; P_{112}(1-P_{112})+n_{x_2}P_{121}\left(1-P_{121}\right)$$
$$+n-n_{x_1}-n_{x_2}-n_{x_1x_2}P_{122}\left(1-P_{122}\right)\}$$
$$+n_{x_1x_2}P_{112}(1-P_{112}-2P_{111})+n_{x_1}\; P_{111}(1+P_{111}-2P_{112})$$
$$+n_{x_2}P_{122}(1-P_{122}-2P_{121})+(n-n_{x_1}-n_{x_2}-n_{x_1x_2})$$
$$P_{121}(1-P_{121}-2P_{122})$$

$$\therefore V\left(\hat{P}_{B_1}\right)=V(P_B)\,n_{x_1x_2}P_{112}\left(1-P_{112}-2P_{111}\right)$$
$$+n_{x_1}P_{111}(1-P_{111}-2P_{112})+n_{x_2}P_{122}(1-P_{122}-2P_{121})$$
$$+(n-n_{x_1}-n_{x_2}-n_{x_1x_2})P_{121}(1-P_{121}-2P_{122})$$

then $V\left(P_{B_1}\right)>V\left(P_B\right)$ if

$P_{111}+2P_{112}<1$ and $P_{112}+2P_{121}<1$

and $P_{112}+2P_{111}<1$ and $P_{121}+2P_{122}<1$

These conditions can be satisfied if

$$3P_{111}+3P_{112}+3P_{121}+3P_{122}<4$$

$$P_{111}+P_{112}+P_{121}+P_{122}<\frac{4}{3}$$

Which is always satisfied

Since $P_{111}+P_{112}+P_{121}+P_{122}<1$

III. Bivariate Ratio Type Median Estimator:

Let $(X_{11}, X_{12}, Y_1),\ldots \left(X_{n_1}, X_{n_2}, Y_n\right)$ be the associated values of the variables X_1, X_2, Y for the units in S_n. When the values of the auxiliary variables X_{i1} , X_{i2} are available a natural modification of the ratio estimator proposed by Kuk and Mak (1988) can be expressed as follows:

Let $M_{YR_1} = M_Y M_{X_1} / M_{X_1}$

and $M_{YR2} = M_Y M_{X_2} / M_{X_2}$

Let $M_{BYR} = W_1 M_Y M_{X_1} / M_{X_1} + W_2 M_Y M_{X_2} / M_{X2}$

Where W_i are weights to be determined to maximize the precision of $\hat{M}_{BYR}$ subject to $\Sigma w_i = 1$

The method is described for two X- varieties since this should be the most frequent application.
Hence,

$$V(M_{BYR}) = W_1^2 V(M_{YR1}) + 2W_1 W_2 Cov(M_{YR1}, M_{YR2}) + W_2^2 V(M_{YR2})$$

Let $V_{11} = V(M_{YR1})$
$V_{12} = Cov(M_{YR1}, M_{YR2})$
$V_{22} = V(M_{YR2})$

Then $V(\hat{M}_{BYR}) = W_1^2 V_{11} + 2W_1 W_2 V_{12} + W_2^2 V_{22}$

Where the values of W_1 , W_2 can be determined by minimizing $V(\hat{M}_{BYR})$ under the condition

$W_1 + W_2 = 1$

It follows that

$$W_1 = \frac{V_{22} - V_{12}}{V_{11} + V_{22} - 2V_{12}} \qquad W_2 = \frac{V_{11} - V_{12}}{V_{11} + V_{22} - 2V_{12}}$$

then

$$M_{min}(M_{BYR}) = \frac{V_{11}V_{22} - V_{12}^2}{V_{11} + V_{22} - 2V_{12}}$$

where

$$V_{11} = n^{-1}(1-f)\left[\frac{1}{4}\{f_T(M_Y)\}^{-2} + \frac{1}{4}\left(\frac{M_Y}{M_{X1}}\right)^2 \{f_{X1}(M_{X1})\}^{-2}\right.$$

$$-2\left(\frac{M_Y}{M_{X1}}\right)\{f_Y(M_Y)f_{X1}(M_{X1})\}^{-1}\left(P_{11}-\frac{1}{4}\right)\Bigg]$$

$$V_{22}=n^{-1}(1-f)\left[\frac{1}{4}\{f_Y(M_Y)\}^{-2}+\frac{1}{4}\left(\frac{M_T}{M_{X2}}\right)^2\{f_{X2}(M_{X2})\}^{-2}\right.$$

$$\left.-2\left(\frac{M_T}{M_{X2}}\right)\{f_Y(M_Y)f_{X2}(M_{X1})\}^{-1}\left(P_{11}-\frac{1}{4}\right)\right]$$

$$V_{12}=Cov(M_{YR1},M_{YR2})$$

$$M_{YR1}-M_Y=(M_{X1}M_Y-M_YM_{X1})/M_{X1}$$

Since $M_{X1}/M_{X1}\to 1$ in probability

$M_{YR1}-M_Y$ has the same asymptotic distribution as

$$(M_{X1}M_Y-M_YM_{X1})/M_{X1}=(M_Y-M_Y)-\left(\frac{M_Y}{M_{X1}}\right)(M_{X1}-M_{X1})$$

In the same way

$M_{YR2}-M_Y$ has the same asymptotic distribution as

$$(M_{X2}M_Y-M_YM_{X2})/M_{X2}=(M_Y-M_Y)-\left(\frac{M_Y}{M_{X2}}\right)(M_{X2}-M_{X2})$$

then

$$Cov(M_{YR1},M_{YR2})$$

$$=Cov\left[\left\{(M_Y-M_Y)-\frac{M_Y}{M_{X1}}(M_{X1}-M_{X1})\right\},\right.$$

$$\left.\left\{(M_Y-M_Y)-\frac{M_Y}{M_{X2}}(M_{X2}-M_{X2})\right\}\right]$$

$$=V(M_Y)-\frac{M_Y}{M_{X2}}Cov(M_Y,M_{X2})-\frac{M_Y}{M_{X1}}Cov(M_Y,M_{X1})$$

$$+\frac{M_Y^2}{M_{X1}M_{X2}}Cov(M_{X1},M_{X2})$$

IV. The Mean Square Error of the Bivariate Ratio Type Median Estimator Compared to that of the Ratio Type Median Estimator

$$V(M_{YR}) - M_{min}(M_{BYR}) =$$

$$V_{11} - \frac{V_{11}V_{22} - V_{12}^2}{V_{11} + V_{22} - 2V_{12}} = \frac{(V_{11} - V_{12})^2}{V_{11} + V_{22} - 2V_{12}}$$

$$M_{min}(M_{BYR}) < M(M_{YR}) \qquad \text{if}$$

$$V_{11} + V_{22} - 2V_{12} > 0$$

$$\frac{1}{2}(V_{11} + V_{22}) > V_{12}$$

This inequality is always satisfied since the equality holds only if the two auxiliary variables are identical, that is:

$$E(M_{YR1} - M_{YR1} - M_{YR2} + M_{YR2})^2 \geq 0$$

$$E[(M_{YR1} - M_{YR1}) - (M_{YR2} - M_{YR2})]^2 \geq 0$$

$$E(M_{YR1} - M_{YR1})^2 + E(M_{YR2} - M_{YR2})^2$$

$$-2E(M_{YR1} - M_{YR1})(M_{YR2} - M_{YR2}) \geq 0$$

$$V(M_{YR1}) + V(M_{YR2}) - 2Cov(M_{YR1}, M_{YR2}) \geq 0$$

$$V_{11} + V_{22} \geq 2V_{12}$$

$$\frac{1}{2}(V_{11} + V_{22}) \geq V_{12}$$

V. Numerical Example:

Consider data for three variables where the dependent variable is water per capita with cubic meters per year in 1996. The first auxiliary variable is the population in millions and the second auxiliary variable is the population growth rate. A simple random sample of 24 countries is chosen from 48 water short countries, table (2) illustrates the three way classification of the sample according to the dependent variable and the two auxiliary variables. Table (3) and Table (4) illustrate the two way classification according to the dependent variable and one of the auxiliary variable at a time. The distribution of the dependent variable is skewed since the skeweness measure is. 413. So, the median estimator is more appropriate than the mean. Some statistical measures are illustrated in table (1) for the population and the sample. The statistics in the table are used to estimate the median water per capita using the two proposed estimators. The estimates and their relative efficiencies are illustrated in tables (6) and (7). Using the ratio type median estimator is not good in this example since the correlation coefficients between the dependent variable and the auxiliary variables are very low. These correlations are illustrated in table (2).

Table (1)
Population and Sample Characteristics of The Dependent and the Auxiliary Variables

	Population			Sample		
	Wpc	Pop	Gr	Wpc	Pop	Gr
Mean	1400.85	36.80	2.10	1371.59	19.24	1.84
Std dev	983.11	133.46	.95	806.77	21.10	.988
Kurtosis	- .787	45.09	-.29	-.47	.25	-.853
Skewness	.413	6.63	-.37	0.43	1.17	-.452
Range	3470	928.70	3.80	2882.0	68.0	3.20
Maximum	3552	929.0	3.90	2964.0	68.40	3.30
Minimum	82.0	.30	.10	82.0	.40	.10
Percentile						
25	457.25	2	1.60	400.25	2.25	1.53
50	1228.0	9.35	2.15	1215	16.15	2.05
75	1965	27.03	2.85	1820.25	38.15	2.65

Table (2)
The Correlation Between the Dependent and the Auxiliary Variables

	Wpc	Pop	Gr
Wpc	1	.1618	.2522
Pop		1	-.0395
Gr			1

Table (3)
Water per capita by population and growth rate

	$X_1 \le MX_1$, $X_1 \le MX_2$	$X_1 \le MX_1$, $X_1 > MX_2$	$X_1 > MX_1$, $X_1 \le MX_2$	$X_1 > MX_1$, $X_1 > MX_2$
$Y \le M_Y$	4 (.1666)	3 (.125)	2 (.083)	2 (.083)
$Y > M_Y$	1 (.042)	1 (.0416)	5 (.208)	6 (.25)
	5 (.20833)	4 (.0416)	7 (.208)	8 (.25)

$\hat{P}_B = .5243804$ $\quad \therefore M_Y = 1425.6$

$V(\hat{P}_B) = .04736$

Table (4)
Water Per Capita by Population

	$X_1 \le MX_1$	$X_1 > MX_1$
$Y \le M_T$	7 (.29166)	4 (.1666)
$Y > M_T$	2 (.08333)	11 (.458333)
	9 (.375)	15 (.625)

$\hat{P}_{B_1} = .4582562$ $M_Y = 1227.46$

$V(\hat{P}_{B_1}) = .0821305$

Table (5)

	$X_2 \le MX_2$	$X_2 > MX_2$
$Y \le M_T$	6 (.25)	5 (.20833)
$Y > M_T$	6 (.25)	7 (.29166)

$\hat{P}_{B_2} = .458333$ $M_Y = 1227.49$

$V(\hat{P}_{B_2}) = .0881057$

Table (6)
Relative Efficiency of The Position Median Estimator

Method of Estimation	Estimate	Stand error	Relative efficiency
Position (first variable)	1227.46	.28652	35.95
Position (second variable)	1337.49	.29683	33.51
Position (Two auxiliary)	1425.6	.21762	6234
Median estimator	1228	1.171829	1.00

Table(7)
Relation Efficiency of the Bivariate Ratio Median estimator

Method of Estimation	Estimates	Relative efficiency
Ratio Median(first variable)	703	1.006
Ratio Median (second variable)	1158	0.740
Ratio Median (two auxiliary variables)	714	1.007
Median estimator	1228	1.00

IV Conclusion:

From table (5) note that using the position median estimator in presence of auxiliary information increased the precision of the estimates than using the dependent variable alone. Using two auxiliary information increased the precision 62 times that using the dependent variable alone.

From table (6) we note also that using the ratio median estimator in presence of auxiliary information increased the precision of the estimates except using the second auxiliary variable alone. Adding auxiliary variables increased the precision although this is not good example for ratio type median estimator. In this example the position estimator works better the comparison between the two approaches is very difficult since the theory behind the two approaches are radically different.

References:

1. ***Bandyopadhyay, S.*** (1980), Improved Ratio and Product Estimators, Sankhya, 2, series c. pt.1 and 2, pp. 45 – 49.

2. ***Cochran, W.G.,*** (1977), Sampling techniques, 3rd edition, John Wiley New York.

3. ***Chang D.S., S.***(1986),The Asymptotic Distribution of multivariate Producer Estimator, Chinese Journal of mathematics, Vol. 14, No.3.

4. ***Hussein,*** **M.A.** (1988), On Increasing the Precision of Finite Population Estimators Using Auxiliary Variables, unpublished ph. D. Thesis University of Lowa USA.

5. ***Hussein, M.A.*** (1992), Alternative Estimators for Stratified Random Sampling The Egyptian Statistical Journal, ISSR Cairo University, Vol 35 No. 2, 1992.

6. ***Hussein M.A.*** (1995), An Estimator in Cluster Sample Combined with Stratification. The Egyptian Statistical Journal, ISSR, Cairo University, Vol 46 No.2,.

7. ***Hussein M.A.*** (1998), Estimator Finite Population Using Stratified Sample in the Presence of the Auxiliary Information. The Egyptian Statistical Journal, ISSR, Cairo University, Vol 15 No.

8. ***Hussein M.A.*** (1999), Separate Combined Ratio Estimator for the Median. The Egyptian Statistical Journal, ISSR, Cairo University, Vol 19 No.

9. ***Ouyang, Z., Srivastaya J.N. and Schreuder, H.T.*** (1992), IKA General Ratio Estimator and Its Application in Model Based Inference, Ann. Inst. Statist. Math. Vol. 45, No. 1,113 - 127.

10. Population Reports, Volume xxv1, Number 1, September 1998.

11. ***Sekkappan R.M.*** (1996), Estimation in Sampling From Finite Populations Under the General Linerar Regression Model. Journal of Indian Statistical association, Vol. 24, 91 – 98.

12. ***Scheaffer, R.L., Mendenhall, W. and Ott, L.*** (1993), Elementary Survey Sampling. Wadsworth, Inc.

13. ***Srivenkataramana, T.*** (1980), Dual to the Ratio Estimator in Sample Surveys, Biometrika, 67, 1. Pp. 193 – 204.