Practical estimators of a population total which reduce the impact of large observations

Jörgen Dalén

SCB
R & D Report
Statistics Sweden
Research - Methods - Development
U/STM - 32

INLEDNING

TILL

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Föregångare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

Efterföljare:

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report, U/STM 1987:32. Practical estimators of a population total which reduce the impact of large observations / Jörgen Dalén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-RnD-USTM-1987-32

Practical estimators of a population total which reduce the impact of large observations

Jörgen Dalén

SCB
R & D Report
Statistics Sweden
Research - Methods - Development
U/STM - 32

Från trycket	March 1987
Producent	Statistiska centralbyrån, Enheten för statistiska metoder
Ansvarig utgivare	Staffan Wahlström
Förfrågningar	Jörgen Dalén, tel 7834494

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1987

ABSTRACT

A class of estimators is presented, which is designed to reduce the impact of large observations in highly-skewed data such as those from business surveys. The class is defined for arbitrary inclusion probabilities. One estimator in this class is shown to have particularly good properties. This estimator and two others are compared and an example from a Swedish financial enterprise survey is presented where the proposed technique has been successfully applied.

Key words:

Finite populations, outliers, skewed data

SCB STATISTICS SWEDEN

1987-03-03

PRACTICAL ESTIMATORS OF A POPULATION TOTAL WHICH REDUCE THE IMPACT OF LARGE OBSERVATIONS

by

Jörgen Dalén

CONTENTS

1. Introduction

Populations with highly-skewed variables are common in survey practice. Well-known examples are production, sales, investment and employment in enterprises or incomes and fortunes of individuals. Surveys of these types of populations and particularly estimates for small domains in them give rise to special difficulties, one of which is treated in this paper, namely the problem of large, true observations with small inclusion probabilities which, according to the Horvitz-Thompson estimator, are to be inflated with the inverse of this probability.

Of course, at the design stage you theoretically should have chosen your inclusion probabilities approximately proportional to the variable values of the survey units, perhaps with a take-all stratum for the very large ones. But in practice, due to insufficient beforehand knowledge and the rapid changes of the units, you often end up with some observations far away from the bulk.

Sometimes this problem is addressed as an outlier problem. Strictly speaking this is not a proper label since the large observation is correct and comes from the same distribution (population) as the rest of the observations. It is also a fact that the vast literature dealing with outliers in statistical data does not address the estimation problem in finite populations.

Instead there are some scattered papers with direct relevance to our problem.

Searls (1963) in his Ph.D. thesis compares several estimators of the population mean, some of which are based on changing the weight or value of those observations, which are larger than a predetermined value t. The other estimators in the comparison use Winsorizing or trimming techniques. Searls (1966) is an accessible short version of his thesis.

Fuller (1970) studied one-sided Winsorized means for skewed populations, Winsorization being applied to the largest observations only, assuming that the right tail of the distribution is well approximated by the tail of a Weibull distribution.

Jenkins, Ringer and Hartley (1973) studied root estimators, that use root transformations of different degrees (mainly square roots) of the original observations. They showed that the square root estimator can provide substantial gains in efficiency over the sample mean when sampling from highly skewed populations.

Ernst (1980) continues Searls work and proves under certain conditions (including a continuous population distribution) that the estimator which substitutes t for all sample observations larger than t has minimum mean square error among several classes of estimators of the mean.

Hidiroglou and Srinath (1981) explicitly studied simple random sampling without replacement from finite populations and analyzed four estimators in which the weights of the large units were changed (decreased). They compared the estimators unconditionally as well as conditionally on the number of large units in the sample.

Moyer and Geissler (1984) use a different technique for identifying the large observations based on the ordered sample and an estimation procedure called semi-Winsorization. They also present a useful bias adjustment procedure designed to reconcile domain estimates with global estimates.

In this paper a class of estimators is presented which technically work by changing the value instead of the weight of the large observations. Within this class three estimators are compared. One estimator is equivalent to changing the weights of the large units to one. Another estimator corresponds to Searls-Ernst estimator which substitutes t for all sample observations larger than t. The third estimator is in a sense a combination of the two others since it gives a weight of one to that part of the value which is larger than t but the unadjusted weight to the value t itself. This estimator in a sense tries to catch the "limit of representativity" by inflating the observed value only up to this limit.

In section 2 the estimators are presented and some (very simple) theory is given concerning their error structure. In section 3 numerical examples are given based on simple random samples from an approximately lognormal finite population. In section 4 an example from a Swedish financial enterprise survey with stratified random sampling is given, where the practical usefulness and applicability of the proposed technique is also demonstrated. In section 5 there is some concluding discussion of the problem.

2. The estimators

We want to estimate a population total $Y = \Sigma_1^N y_k$. The class, in which the estimators of this paper falls, is as follows:

$$\hat{Z} = \Sigma_S z_k/\pi_k \quad \text{where } z_k = \begin{cases} f(y_k) \text{ if } y_k/\pi_k > T \\ g(y_k) \text{ if } y_k/\pi_k \leq T \end{cases}$$

y_k is the variable value and π_k the inclusion probability of unit k. Σ_S stands for summation over the sample. For the estimators studied in this paper $f(y_k)<y_k$ and $g(y_k)=y_k$ since we deal with positively skewed populations. (For negatively skewed populations we would choose $f(y_k)=y_k$ and $g(y_k)<y_k$.)

Taking expectations and variances of $\hat{Z}$ we get

$$E(\hat{Z}) = \Sigma_1^N z_k,$$

$$V(\hat{Z}) = \Sigma_1^N \Sigma_1^N (\pi_{kl}-\pi_k\pi_l)(z_k/\pi_k)(z_l/\pi_l) \text{ and}$$

$$\text{Bias}(\hat{Z}) = \Sigma_1^N (z_k-y_k)$$

The results are obvious regarding z_k as just any kind of variable in the survey. As unbiased estimates of $V(\hat{Z})$ and Bias $(\hat{Z})$ we, by the same reasoning, obtain:

$$\hat{V}(\hat{Z}) = \Sigma_S \Sigma_S (1-\pi_k\pi_l/\pi_{kl})(z_k/\pi_k)(z_l/\pi_l) \text{ and}$$

$$\hat{\text{Bias}}(\hat{Z}) = \Sigma_S(z_k-y_k)/\pi_k$$

and we may estimate the mean square error of $\hat{Z}$ as

$$\hat{\text{MSE}}(\hat{Z}) = \hat{V}(\hat{Z}) + \{\hat{\text{Bias}}(\hat{Z})\}^2$$

The three estimators that we consider in this paper have the following form:

$\hat{Z}_1 = \Sigma_{S_0} y_k + \Sigma_{S_1} y_k/\pi_k$ with $f(y_k) = \pi_k y_k$,

$\hat{Z}_2 = \Sigma_{S_0} T + \Sigma_{S_1} y_k/\pi_k$ with $f(y_k) = \pi_k T$ and

$\hat{Z}_3 = \Sigma_{S_0} (T+y_k-\pi_k T) + \Sigma_{S_1} y_k/\pi_k$ with $f(y_k) = \pi_k(T+y_k-\pi_k T)$

Here $S_0 = \{k\varepsilon S: y_k/\pi_k > T\}$ and $S_1 = S-S_0$.

In the sequel these three estimators will be compared, using the mean square error criterion, to the traditional unbiased HT-estimator $\hat{Z}_0 = \Sigma_S \, y_k/\pi_k$ (also a member of the $\hat{Z}$-class with $f(y_k) = y_k$)

In Diagram 1 a geometrical interpretation of the three estimators is given with the observed value at the x-axis and the inflated value at the y-axis. We observe that a bad property of $\hat{Z}_1$ is its being discontinuous at T, which means that in a certain region an increase of the observed value leads to a smaller estimate. $\hat{Z}_2$ has another bad property. For sufficiently large observed values its "inflated" value becomes smaller than the observed value which is absurd. The estimator $\hat{Z}_3$ lacks these negative aspects in that it is continuous at T and always increasing.

3. A study of lognormal populations

In order to study the properties of these estimators more concretely populations based on fixed percentiles of the lognormal distribution were created. Four populations of size 1000 with different degrees of skewness were made up by choosing different values of σ in its density function

$$f(x) = \exp\left(-\log^2 x/2\sigma^2\right)/\sigma x\sqrt{2\pi}, \text{ giving the skewness}$$

$$G=\{\exp(\sigma^2)+2\}\{\exp(\sigma^2)-1\}^{1/2}$$

For eight different sample sizes (simple random sampling without replacement was used) and nine to thirteen expansion limits the mean square errors of $\hat{Z}_1$, $\hat{Z}_2$ and $\hat{Z}_3$ were calculated relative to the variance of $\hat{Z}_0$. The programming language was SIMULA and the IMSL procedure MDNRIS was used. The results are in table 1 and 2.

We can see from table 1a-d that there is typically a unique optimum value of t ($=\pi_k T=nT/N$ in this case) for which the MSE is smallest for a given population, estimator and sample size. In the table we can observe the approximate optimum point and optimum value (marked with *). In table 2 only the optima are tabulated showing the maximum gain in efficiency that could be obtained in different situations.

We see that the smaller the sample and the skewer the population the more there is to gain by introducing an expansion limit. But in every case something could be gained. The optimum is quite flat which is nice since it is difficult to hit at it exactly. Missing it by a factor of 2 or 3 often does not make much difference. Of course, a grossly improper choice of expansion limit could lead to a larger error than no limit at all. This risk is greater for large samples and non-skewed populations.

Comparing the three estimators we see that $\hat{Z}_1$ performs least well: there is less to be gained in the optima and more to be lost if the expansion limit is poorly chosen. The performance of $\hat{Z}_2$ and $\hat{Z}_3$ is about even with a small preference for $\hat{Z}_3$, especially for large sample sizes, due to its greater robustness against non-optimal choices of the expansion limit. On the other hand the optimum MSE for $\hat{Z}_2$ is usually slightly smaller than that of $\hat{Z}_3$.

Optimally a larger share of the units should be above the expansion limit for small samples and for more skewed populations. For these populations at most some 10% of the population units should be considered large but for large samples less than 1%.

4. An example - exports and imports of services

In this section we give an example, where the estimators presented above are used. This is the Swedish Survey on Exports and Imports of Services.

This survey is carried out annually and its objective is to estimate the total amount of exports as well as imports of services by enterprises in Sweden. Stratified random sampling with about 15 strata is used with inclusion probabilities ranging from about 0.01 to 1 (take-all-strata).

Data from five consecutive years - 1980-84 - were used and the bias, variance and MSE of the three estimators were estimated for nine different expansion limits. The take-all strata were excluded from the calculations. The results are presented in tables 3 and 4. We see that the results for single years are widely apart. The MSE-estimation procedure is not very robust. But, since the population distributions are believed to be quite similar from year to year, we may make the assumption that all the five samples (which are of about the same size) come from a common population and estimate the variance and bias of the estimator by the mean of the five yearly estimates. The MSE is then estimated by the mean of the variance estimates plus the square of the mean of the bias estimates.

These kinds of estimates provide a basis for a choice of estimator and expansion limit in successive surveys, where, as often is the case, the population distribution, the sample size and the allocation of the sample are relatively stable. Of course, there must be a continuous follow-up in order to check that the structure of the population does not change significantly.

In this case we notice that the conclusions from the study of the lognormal population seem to hold here too. $\hat{Z}_1$ is less robust to non-optimal expansion limits and $\hat{Z}_3$ seems to be a little more robust than $\hat{Z}_2$. In the optimum there is not much different, however. Significant improvements in the MSE could be made by introducing an expansion limit also when the optimum point is missed by a factor of two or three.

5. Summary and conclusions

We have presented three estimators which are easy to implement and could reduce the mean square error by a significant amount when large π-expanded observations are present in the population. The studies of lognormal populations and a real survey suggest that $\hat{Z}_3$ is the estimator to be preferred due to its greater robustness to the choice of expansion limit. It also has an attractive property in that it is strictly increasing when the observed values are increasing. $\hat{Z}_1$ and $\hat{Z}_2$ both lack this desirable property. $\hat{Z}_1$, which is the estimator that reduces the weights of the large observations to one, is clearly inferior to the others.

It should further be noticed that it is very simple to estimate the MSE by adding the variance estimate of section 2 to a squared bias estimate. Since a single year bias estimate is very sensitive to large observations it is usually better to estimate the bias as a mean of those in several consecutive surveys.

6 References:

Ernst, L.R. (1980), Comparison of Estimators of the Mean which Adjust for Large Observations, Sankhya, 42, Ser. C, 1-16.

Fuller, W.A. (1970), Simple Estimators for the Mean of Skewed Populations, Technical Report prepared for the U.S. Bureau of the Census, Iowa State University, Dept. of Statistics.

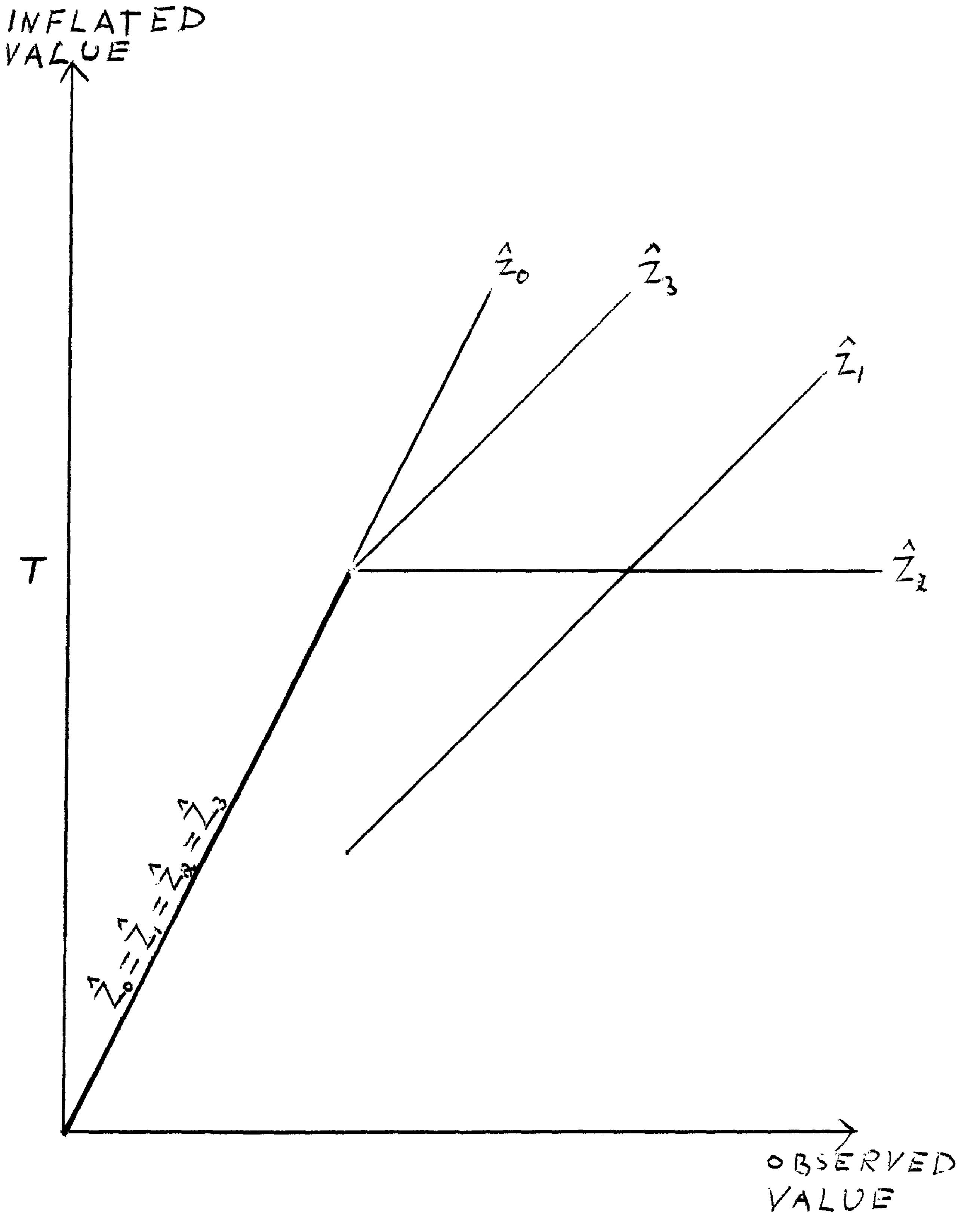
Hidiroglou, M.A. and Srinath, K.P. (1981), Some Estimators of a Population Total from Simple Random Samples Containing Large Units, JASA, 76, 690-695.

Jenkins. O.C., Ringer, L.J. and Hartley, H.O. (1973), Root Estimators, JASA, 68, 414-419.

Moyer, L.M. and Geissler, P.H. (1984), Improving Stability of Annual State Waterfowl Harvest Estimates in Highly-skewed Data, Proceedings of the Section on Survey Research Methods, American Statistical Association, 467-472.

Searls, D.T. (1963), On the 'Large' Observation Problem, unpublished Ph.D. thesis, North Carolina State University.

Searls, D.T. (1966), An Estimator for a Population Mean which Reduces the Effect of Large True Observations, JASA, 61, 1200-1204



<u>DIAGRAM 1</u>: A GEOMETRICAL INTERPRETATION OF THE ESTIMATORS.

Table 1: Lognormal populations. Relative RMSE $\{\sqrt{MSE(\hat{Z}_i)}/\sqrt{V(\hat{Z}_0)}\}$ of the examined estimators for different sample sizes and expansion limits.

Table 1a: $G=4.505, y_{MAX}=26.9$

		t=									$\sqrt{V(\hat{Z}_0)}$
		1	2	3	5	7	10	12.5	16	20	
Units larger than t, o/oo		500	244	136	54	26	11	6	3	1	
n=5	$\hat{Z}_1$	1.482	1.119	0.882	0.693	0.686*	0.749	0.803	0.858	0.925	
	$\hat{Z}_2$	0.957	0.652	0.587*	0.660	0.750	0.842	0.891	0.934	0.965	933
	$\hat{Z}_3$	0.953	0.650	0.587*	0.661	0.751	0.843	0.891	0.935	0.965	
n=10	$\hat{Z}_1$	2.085	1.555	1.191	0.838	0.747*	0.767	0.810	0.861	0.926	
	$\hat{Z}_2$	1.349	0.867	0.696	0.686*	0.758	0.844	0.891	0.934	0.965	658
	$\hat{Z}_3$	1.337	0.861	0.694	0.687*	0.759	0.845	0.892	0.935	0.966	
n=20	$\hat{Z}_1$	2.931	2.172	1.638	1.067	0.856	0.802*	0.824	0.866	0.926	
	$\hat{Z}_2$	1.913	1.190	0.878	0.737*	0.773	0.847	0.892	0.935	0.965	463
	$\hat{Z}_3$	1.875	1.170	0.868	0.737*	0.775	0.849	0.894	0.936	0.966	
n=50	$\hat{Z}_1$	4.559	3.366	2.514	1.547	1.110	0.896	0.864*	0.880	0.929	
	$\hat{Z}_2$	3.067	1.869	1.289	0.876	0.818*	0.857	0.896	0.936	0.966	288
	$\hat{Z}_3$	2.915	1.782	1.239	0.864	0.819*	0.861	0.899	0.938	0.967	
n=100	$\hat{Z}_1$	6.274	4.627	3.444	2.075	1.414	1.023	0.923	0.901*	0.933	
	$\hat{Z}_2$	4.453	2.695	1.811	1.084	0.895	0.876*	0.902	0.937	0.966	199
	$\hat{Z}_3$	4.011	2.435	1.652	1.029	0.884	0.880*	0.908	0.942	0.969	
n=200	$\hat{Z}_1$	8.366	6.166	4.583	2.733	1.808	1.206	1.015	0.938*	0.940	
	$\hat{Z}_2$	6.678	4.026	2.669	1.463	1.056	0.918	0.917*	0.941	0.967	132
	$\hat{Z}_3$	5.348	3.238	2.170	1.259	0.985	0.914*	0.924	0.949	0.972	
n=300	$\hat{Z}_1$	9.586	7.065	5.250	3.123	2.048	1.327	1.081	0.968	0.948*	
	$\hat{Z}_2$	8.742	5.266	3.474	1.840	1.232	0.969	0.935*	0.946	0.968	101
	$\hat{Z}_3$	6.129	3.710	2.479	1.406	1.057	0.941	0.939*	0.957	0.976	
n=500	$\hat{Z}_1$	10.465	7.717	5.737	3.416	2.239	1.436	1.151	1.007	0.963*	
	$\hat{Z}_2$	13.353	8.035	5.280	2.714	1.675	1.118	0.992	0.961*	0.971	66
	Z_3	6.698	4.062	2.719	1.537	1.135	0.981	0.964*	0.970	0.983	

Table 1b: G=9.205, y_{MAX}=139

		t=													$\sqrt{V(\hat{Z}_0)}$
		1	5	7	10	15	20	25	30	40	49	60	70	100	
Units larger than t, o/oo		500	142	97	62	36	23	16	12	7	5	3	2	1	
n=5	$\hat{Z}_1$	0.832	0.610	0.550	0.500	0.477*	0.489	0.514	0.542	0.602	0.643	0.705	0.752	0.826	
	$\hat{Z}_2$	0.690	0.446	0.428*	0.442	0.497	0.554	0.604	0.647	0.717	0.765	0.812	0.847	0.918	3399
	$\hat{Z}_3$	0.687	0.445	0.428*	0.443	0.498	0.555	0.605	0.685	0.718	0.766	0.813	0.848	0.918	
n=10	$\hat{Z}_1$	1.173	0.846	0.750	0.656	0.582	0.557*	0.560	0.574	0.619	0.654	0.710	0.755	0.827	
	$\hat{Z}_2$	0.977	0.590	0.533	0.507*	0.528	0.570	0.613	0.653	0.719	0.767	0.813	0.847	0.918	2397
	$\hat{Z}_3$	0.968	0.586	0.531	0.507*	0.529	0.572	0.615	0.655	0.721	0.768	0.814	0.849	0.919	
n=20	$\hat{Z}_1$	1.651	1.181	1.036	0.886	0.746	0.672	0.642	0.634*	0.650	0.674	0.720	0.761	0.829	
	$\hat{Z}_2$	1.388	0.807	0.698	0.618	0.586*	0.602	0.632	0.665	0.725	0.770	0.814	0.848	0.918	1686
	$\hat{Z}_3$	1.361	0.794	0.690	0.614	0.586*	0.604	0.635	0.668	0.728	0.773	0.817	0.851	0.920	
n=50	$\hat{Z}_1$	2.569	1.830	1.596	1.344	1.088	0.927	0.832	0.780	0.732	0.728*	0.748	0.777	0.834	
	$\hat{Z}_2$	2.228	1.264	1.061	0.879	0.738	0.693	0.688*	0.701	0.741	0.778	0.819	0.851	0.919	1050
	$\hat{Z}_3$	2.118	1.207	1.017	0.850	0.726	0.690*	0.691	0.706	0.747	0.785	0.825	0.856	0.922	
n=100	$\hat{Z}_1$	3.537	2.516	2.190	1.835	1.464	1.218	1.061	0.963	0.844	0.807	0.791*	0.802	0.844	
	$\hat{Z}_2$	3.238	1.821	1.511	1.217	0.955	0.833	0.780	0.761*	0.769	0.794	0.826	0.855	0.920	723
	$\hat{Z}_3$	2.916	1.649	1.375	1.120	0.900	0.806	0.769	0.761*	0.778	0.805	0.838	0.865	0.926	
n=200	$\hat{Z}_1$	4.719	3.355	2.918	2.441	1.934	1.589	1.360	1.211	1.007	0.927	0.861	0.846*	0.863	
	$\hat{Z}_2$	4.856	2.721	2.243	1.778	1.335	1.096	0.963	0.889	0.833	0.829*	0.844	0.865	0.922	482
	$\hat{Z}_3$	3.891	2.195	1.821	1.464	1.136	0.972	0.889	0.850	0.830*	0.840	0.862	0.883	0.935	
n=300	$\hat{Z}_1$	5.410	3.848	3.347	2.798	2.215	1.815	1.546	1.369	1.118	1.012	0.916	0.883	0.881*	
	$\hat{Z}_2$	6.358	3.557	2.927	2.308	1.703	1.362	1.156	1.030	0.908	0.872	0.867*	0.878	0.924	368
	$\hat{Z}_3$	4.461	2.518	2.089	1.674	1.287	1.085	0.976	0.917	0.874	0.872*	0.884	0.900	0.943	
n=500	$\hat{Z}_1$	5.914	4.214	3.669	3.073	2.439	2.002	1.708	1.512	1.231	1.109	0.989	0.940	0.917*	
	$\hat{Z}_2$	9.712	5.427	4.459	3.499	2.546	1.986	1.628	1.389	1.116	0.998	0.936	0.917*	0.932	241
	$\hat{Z}_3$	4.882	2.770	2.304	1.853	1.430	1.204	1.077	1.004	0.939	0.923*	0.924	0.932	0.960	

1c: G=14.697, y_{MAX}=721

		t=												$\sqrt{\hat{V}(Z_0)}$
		5	10	15	20	30	40	50	70	100	200	300	500	
Units larger than t, o/oo		211	125	88	67	45	33	25	17	11	4	2	1	
n=5	$\hat{Z}_1$	0.440	0.401	0.376	0.358	0.338	0.330*	0.331	0.345	0.376	0.496	0.598	0.702	
	$\hat{Z}_2$	0.371	0.328	0.309	0.304*	0.312	0.332	0.355	0.403	0.468	0.623	0.725	0.856	14119
	$\hat{Z}_3$	0.370	0.327	0.309	0.304*	0.313	0.333	0.356	0.404	0.469	0.624	0.726	0.857	
n=10	$\hat{Z}_1$	0.620	0.563	0.522	0.491	0.451	0.427	0.412	0.405*	0.416	0.509	0.603	0.704	
	$\hat{Z}_2$	0.523	0.452	0.413	0.391	0.375*	0.378	0.390	0.424	0.479	0.625	0.726	0.857	9959
	$\hat{Z}_3$	0.518	0.448	0.411	0.390	0.375*	0.379	0.391	0.426	0.481	0.627	0.728	0.858	
n=20	$\hat{Z}_1$	0.871	0.789	0.729	0.682	0.617	0.572	0.538	0.503	0.486*	0.533	0.613	0.707	
	$\hat{Z}_2$	0.741	0.633	0.568	0.526	0.478	0.457	0.452*	0.463	0.501	0.630	0.727	0.857	7006
	$\hat{Z}_3$	0.727	0.622	0.560	0.519	0.474	0.456	0.452*	0.466	0.505	0.635	0.731	0.859	
n=50	$\hat{Z}_1$	1.356	1.226	1.130	1.053	0.943	0.862	0.795	0.715	0.646	0.596*	0.640	0.718	
	$\hat{Z}_2$	1.187	1.007	0.894	0.814	0.709	0.646	0.606	0.569	0.563*	0.646	0.733	0.858	4363
	$\hat{Z}_3$	1.129	0.960	0.855	0.781	0.685	0.628	0.595	0.566*	0.567	0.656	0.742	0.864	
n=100	$\hat{Z}_1$	1.868	1.688	1.554	1.448	1.292	1.175	1.077	0.952	0.835	0.684	0.682*	0.736	
	$\hat{Z}_2$	1.723	1.459	1.290	1.168	1.000	0.891	0.815	0.723	0.663*	0.674	0.743	0.859	3003
	$\hat{Z}_3$	1.555	1.320	1.170	1.063	0.919	0.827	0.765	0.693	0.654*	0.689	0.760	0.872	
n=200	$\hat{Z}_1$	2.494	2.255	2.076	1.933	1.724	1.565	1.430	1.255	1.083	0.817	0.754*	0.769	
	$\hat{Z}_2$	2.584	2.186	1.928	1.740	1.478	1.299	1.170	0.997	0.853	0.736*	0.766	0.863	2002
	$\hat{Z}_3$	2.078	1.764	1.562	1.417	1.216	1.084	0.990	0.872	0.785	0.748*	0.794	0.887	
n=300	$\hat{Z}_1$	2.864	2.590	2.386	2.222	1.982	1.800	1.644	1.442	1.241	0.912	0.813	0.802*	
	$\hat{Z}_2$	3.383	2.860	2.521	2.273	1.925	1.685	1.508	1.265	1.048	0.809	0.794*	0.868	1529
	$\hat{Z}_3$	2.389	2.029	1.798	1.631	1.400	1.246	1.136	0.992	0.879	0.798*	0.825	0.902	
n=500	$\hat{Z}_1$	3.148	2.851	2.630	2.453	2.194	1.997	1.829	1.611	1.393	1.027	0.901	0.863*	
	$\hat{Z}_2$	5.167	4.367	3.847	3.466	2.927	2.554	2.275	1.881	1.512	1.007	0.879*	0.884	1001
	$\hat{Z}_3$	2.633	2.245	1.996	1.816	1.568	1.401	1.282	1.125	0.997	0.879*	0.882	0.931	

Table 1d: G=23.077, y_{MAX}=19.365

		t=												$\sqrt{V(\hat{Z}_0)}$
		10	20	50	100	200	300	500	1000	2000	3000	5000	10000	
Units larger than t, o/oo		221	159	96	62	39	29	19	11	6	4	2	1	
n=5	$\hat{Z}_1$	0.212	0.209	0.203	0.196	0.188	0.184	0.180*	0.186	0.217	0.255	0.352	0.481	
	$\hat{Z}_2$	0.206	0.200	0.190	0.180	0.173	0.172*	0.180	0.217	0.295	0.362	0.470	0.658	309996
	$\hat{Z}_3$	0.205	0.200	0.189	0.180	0.173*	0.173	0.181	0.219	0.297	0.364	0.472	0.660	
n=10	$\hat{Z}_1$	0.299	0.295	0.287	0.276	0.263	0.255	0.244	0.237*	0.251	0.278	0.362	0.485	
	$\hat{Z}_2$	0.292	0.284	0.268	0.253	0.336	0.229	0.226*	0.245	0.308	0.369	0.473	0.659	218649
	$\hat{Z}_3$	0.290	0.281	0.266	0.251	0.235	0.228	0.226*	0.247	0.311	0.372	0.477	0.662	
n=20	$\hat{Z}_1$	0.422	0.416	0.403	0.388	0.369	0.355	0.335	0.314	0.306*	0.319	0.382	0.493	
	$\hat{Z}_2$	0.415	0.403	0.380	0.357	0.330	0.314	0.297	0.293*	0.332	0.383	0.480	0.660	153825
	$\hat{Z}_3$	0.408	0.396	0.374	0.351	0.325	0.310	0.295*	0.295	0.338	0.390	0.486	0.666	
n=50	$\hat{Z}_1$	0.657	0.649	0.629	0.605	0.574	0.551	0.516	0.470	0.429	0.415*	0.434	0.516	
	$\hat{Z}_2$	0.667	0.648	0.610	0.571	0.523	0.492	0.452	0.408	0.399*	0.425	0.499	0.665	95787
	$\hat{Z}_3$	0.636	0.618	0.582	0.546	0.502	0.474	0.439	0.403*	0.406	0.437	0.514	0.679	
n=100	$\hat{Z}_1$	0.908	0.896	0.869	0.836	0.793	0.761	0.711	0.642	0.571	0.533	0.508*	0.552	
	$\hat{Z}_2$	0.969	0.941	0.886	0.829	0.757	0.709	0.645	0.559	0.499	0.492*	0.532	0.673	65925
	$\hat{Z}_3$	0.878	0.853	0.804	0.754	0.692	0.651	0.598	0.531	0.496*	0.504	0.556	0.699	
n=200	$\hat{Z}_1$	1.220	1.204	1.168	1.124	1.067	1.024	0.957	0.863	0.759	0.696	0.623	0.618*	
	$\hat{Z}_2$	1.454	1.412	1.329	1.243	1.134	1.060	0.958	0.812	0.678	0.623	0.602*	0.691	43950
	$\hat{Z}_3$	1.180	1.147	1.083	1.016	0.933	0.877	0.803	0.704	0.629	0.611*	0.630	0.739	
n=300	$\hat{Z}_1$	1.411	1.393	1.352	1.303	1.237	1.189	1.113	1.006	0.885	0.810	0.711	0.678*	
	$\hat{Z}_2$	1.903	1.848	1.740	1.627	1.483	1.386	1.251	1.052	0.855	0.759	0.683*	0.714	33567
	$\hat{Z}_3$	1.366	1.328	1.255	1.179	1.085	1.022	0.938	0.822	0.727	0.695	0.694*	0.776	
n=500	$\hat{Z}_1$	1.585	1.566	1.522	1.471	1.402	1.351	1.271	1.159	1.032	0.952	0.840	0.783*	
	$\hat{Z}_2$	2.907	2.823	2.658	2.484	2.264	2.114	1.905	1.592	1.265	1.085	0.893	0.784*	21975
	$\hat{Z}_3$	1.537	1.498	1.421	1.341	1.242	1.177	1.089	0.967	0.863	0.822	0.803*	0.846	

Table 2: Approximate efficiency gain (relative RMSE) in optimum points for lognormal populations.

Estimator	Sample size	Skewness (G) 4.5	9.2	14.7	23.1
$\hat{Z}_1$	5	.69	.48	.33	.18
	10	.75	.56	.41	.24
	20	.80	.63	.49	.31
	50	.86	.73	.60	.42
	100	.90	.79	.68	.51
	200	.94	.85	.75	.62
	300	.95	.88	.80	.68
	500	.96	.92	.86	.78
$\hat{Z}_2$	5	.59	.43	.30	.17
	10	.69	.51	.38	.23
	20	.74	.59	.45	.29
	50	.82	.69	.56	.40
	100	.88	.76	.66	.49
	200	.92	.83	.74	.60
	300	.94	.87	.79	.68
	500	.96	.92	.88	.78
$\hat{Z}_3$	5	.59	.43	.30	.17
	10	.69	.51	.38	.23
	20	.74	.59	.45	.30
	50	.82	.69	.57	.40
	100	.88	.76	.65	.50
	200	.91	.83	.75	.61
	300	.94	.87	.80	.69
	500	.96	.92	.88	.80

Table 3:
Survey on Exports and Imports of Services, 1980-1984
RMSE (mkr = millions of Sw kr)

Expansion limit (mkr)	1980			1981			1982			1983			1984			Estimate 80-84		
	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$
50	1739	1299	1155	2946	2876	2372	2183	1557	1495	2107	1876	1606	2757	1774	1587	2447	1875	1645
100	1411	814	779	2407	2460	2016	1701	1017	985	1844	1267	1080	1642	1082	987	1802	1319	1168
150	737	633	632	2287	2243	1847	1567	728	714	1413	986	859	1363	876	819	1467	1067	962
200	737	610	609	2287	2061	1715	973	629	626	1247	821	746	1363	737	714	1313	924	855
300	737	586	585	2127	1791	1527	807	642	643	1247	703	703	891	687	690	1147	791	773
400	737	594	595	1863	1628	1434	669	669	669	718	724	728	738	703	710	851	762	770
500	737	634	635	1588	1532	1389	669	669	669	733	733	733	738	744	746	802	765	783
1000	734	734	734	1588	1354	1368	669	669	669	733	733	733	759	759	759	774	821	852
2000 ($\hat{z}_0$)	734	734	734	1724	1724	1724	669	669	669	733	733	733	759	759	759	1007	1007	1007
Max RMSE-decrease	0	20	20	8	>21	>20	0	6	6	2	4	4	3	7	6	>23	24	24

Table 4:
Survey on Exports and Imports of Services
RMSE (mkr = millions of Sw kr)

Expansion limit (mkr)	1980			1981			1982			1983			1984			Estimate 80-84		
	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$
50	1173	1420	942	848	589	530	2444	1997	1892	3125	3386	2732	1197	821	684	1757	1636	1351
100	1024	1156	773	692	375	361	1962	1594	1538	2990	2990	2408	622	523	469	1451	1300	1085
150	1024	982	679	431	344	346	1796	1439	1397	2877	2676	2165	548	452	433	1317	1124	955
200	736	850	628	360	361	362	1796	1311	1278	2706	2403	1960	548	427	425	1191	999	871
300	736	716	608	368	368	368	1604	1146	1125	2706	1950	1648	469	452	454	1102	853	792
400	736	636	613	368	368	368	1604	1038	1027	1982	1705	1514	476	476	476	934	787	770
500	736	631	644	368	368	368	1604	981	978	1684	1628	1481	476	476	476	888	768	775
1000	665	665	665	368	368	368	1171	1171	1171	1684	1447	1462	476	476	476	805	855	888
2000 ($\hat{z}_0$)	665	665	665	368	368	368	1171	1171	1171	1828	1828	1828	476	476	476	1050	1050	1050
Max RMSE-decrease	0	5	8	2	7	6	0	>16	>16	8	>21	>20	2	10	11	>23	27	27

Tidigare nummer av Promemorior från U/STM:

NR

1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)

2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)

3 En presentation av Box-Jenkins metod för analys och prognos av tidsserier. Åke Holmén (1979-12-20)

4 Handledning i AID-analys. Anders Norberg (1980-10-22)

5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)

6 Metoder för evalvering av noggrannheten i SCBs statistik. En översikt. Jörgen Dalén (1981-03-02)

7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)

8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)

9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan Wretman och Göran Örnberg (1983-06-20)

10 Current survey research at Statistics Sweden. Lars Lyberg, Bengt Swensson och Jan Håkan Wretman (1983-09-01)

11 Utjämningsmetoder vid nivåkorrigering av tidsserier med tillämpning på nationalräkenskapsdata. Lars-Otto Sjöberg (1984-01-11)

12 Regressionsanalys för f d statistikstuderande. Harry Lütjohann (1984-02-01)

13 Estimating Gini and Entropy inequality parameters. Fredrik Nygård och Arne Sandström (1985-01-09)

14 Income inequality measures based on sample surveys. Fredrik Nygård och Arne Sandström (1985-05-20)

15 Granskning och evalvering av surveymodeller, tiden före 1960. Gösta Forsman (1985-05-30)

16 Variance estimators of the Gini coefficient - simple random sampling. Arne Sandström, Jan Wretman och Bertil Waldén (Memo, Februari 1985)

17 Variance estimators of the Gini coefficient - probability sampling. Arne Sandström, Jan Wretman och Bertil Waldén (1985-07-05)

18 Reconciling tables and margins using least-squares. Harry Lütjohann (1985-08-01)

19 Ersättningens och uppgiftslämnarbördans betydelse för kvaliteten i undersökningarna om hushåållens utgifter. Håkan L. Lindström (1985-11-29)

20 A general view of estimation for two phases of selection. Carl-Erik Särndal och Bengt Swensson (1985-12-05)

21 On the use of automated coding at Statistics Sweden. Lars Lyberg (1986-01-16)

22 Quality Control of Coding Operations at Statistics Sweden. Lars Lyberg (1986-03-20)

23 A General View of Nonresponse Bias in Some Sample Surveys of the Swedish Population. Håkan L Lindström (1986-05-16)

24 Nonresponse rates in 1970 - 1985 in surveys of Individuals and Households. Håkan L. Lindström och Pat Dean (1986-06-06)

25 Two Evaluation Studies of Small Area Estimation Methods: The Case of Estimating Population Characteristics in Swedish Municipalities for the Intercensal Period. Sixten Lundström (1986-10-14)

26 A Survey Practitioner's Notion of Nonresponse. Richard Platek (1986-10-20)

27 Factors to be Considered in Developing a Reinterview Program and Interviewer Debriefings at SCB. Dawn D. Nelson (1986-10-24)

28 Utredning kring bortfallet i samband med den s k Metropolit-debatten. Sixten Lundström (1986-11-10)

29 Om HINK-undersökningens design- och allokeringsproblematik. Bengt Rosén (1986-10-06)

30 Om HINK-undersökningens estimationsförfarande. Bengt Rosén (1987-01-28)

31 Program för minskat bortfall i SCBs individ- och hushållsundersökningar - förslag från aktionsgruppen för uppgiftslämnarfrågor. Lars Lundgren (1987-02-20)

Kvarvarande exemplar av ovanstående promemorior kan rekvireras från Elseliv Lindfors, U/STM, SCB, 115 81 Stockholm, eller per telefon 08 7834178