International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-7,July 2016
ISSN: 2395-3470
www.ijseas.com

The Grouping Methods and Rank Estimator, Based on Ranked Set sampling, for the linear Error in Variable Models

Ahmed A. M. Al-Radaideh[1]
[1] Colleague of General Studies, Ajman University of Science and technology,
AL Fujairah, UAE
a.alradaideh@ajman.ac.ae

Abstract

The methods of estimating the parameters of Error-In-Variables (EIV) models were extremely discussed in the literature. Wald (1940), Bartlett (1942) and others started the discussion in this area as well as Madansky (1959). In this paper the ranked set sampling method is used to choose a sample and analyze it using EIV model in order to estimate the parameters. The estimations are made by using the grouping methods and the ranked estimator. Moreover a simulation study has been applied in order to compare these methods.

Keywords: *Errors–in-Variables, Grouping Methods, Ranked Set Sampling, Ranked Set Sampling, Simple Random Sampling.*

1. Introduction

Statistical methods are useful tools for assessing the relationships between continuous variables collected for the applied studies. However, it is also important to distinguish these statistical methods. While they are similar mathematically, their purposes are different. Regression analysis has undoubtedly been one of the most widely used techniques in applied statistics. It is well known that the simple linear regression model assumes that the errors can happened just in the explanatory variable but practically, errors can happened in both variables. Here we are interested in the measurement error model where both the response and the explanatory variable are subject to error, consider two variables X and Y has the simple linear relationship:

$$\mathbf{y} = \alpha + \beta\mathbf{x} + \boldsymbol{\varepsilon} \qquad (1)$$

where y is $n \times 1$ vector and $\mathbf{x}$ is $n \times 1$ vector, $\mathbf{x}$ could be random or functional and $\boldsymbol{\varepsilon}$ is an error term of size $n \times 1$, with mean zero and constant variance σ_j^2 such that $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \,\forall\, i \neq j$ and $\text{cov}(\boldsymbol{\varepsilon}, \mathbf{x}) = 0$

Moreover, the unknown parameters of this relationship are α and β denoted by intercept and slope respectively. The details of this model can be found in many text books s. t. (Darper and Smith, 1980; Montegomry and Pech, 1981; Neier et al, 1996).

More general, Chen and Vanness (1999), extend the simple linear relationship to the so-called Measurement Error Model (ME) or error- in- variable model 'EIV'. In this models, consider two latent variables $\boldsymbol{\eta}$ and ξ with the following relationship

$$\boldsymbol{\eta} = \alpha + \beta.\boldsymbol{\xi} \qquad (2)$$

Both variable can't be measure directly instead one use manifest variable Y and X to measure latent variables with error

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad (3)$$
$$\mathbf{x} = \boldsymbol{\xi} + \boldsymbol{\delta} \qquad (4)$$

The model can be simplified as

$$\mathbf{y} = \alpha + \beta(\mathbf{x} - \boldsymbol{\delta}) + \boldsymbol{\varepsilon} \qquad (5)$$

Nothing that $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are the error terms, have the following assumptions; Kendall and Stuart (1979);

1. $E(\varepsilon) = E(\delta) = 0$
2. $\text{cov}(\varepsilon_i, \varepsilon_j) = \text{cov}(\delta_i, \delta_j) = 0 \,\forall i \neq j$
3. $\text{var}(\varepsilon) = \sigma_\varepsilon^2$, $\text{var}(\delta) = \sigma_\delta^2$

There are many methodologies to estimate the unknown parameters α and β for model (3)-(5) suggested in the literatures, one of the most popular is the grouping methods which was proposed by Wald (1940) and Bartlett (1949), alternative methods of estimation of EIV model have been suggested by Theil (1950), Madansky (1959), Kendall and Stuart (1979), Fuller (1987), Chen and Vanness (1999), and Al-Nasser et al (2005).all the proposed estimation methods was done based on selecting Simple Random Sample from the population of the study, which is one of the most common techniques for obtaining such data. Other more structured sampling designs, such as stratified sampling or probability sampling, are also available to help make sure that the obtained data collection provides a good representation of the population of interest. Any such additional structure of this type revolves around how the sample data themselves should be collected in order to provide information's of the larger population. With any of these methods, once the sample items have been chosen the desired measurement(s) is collected from each of the selected items.

The concept of ranked set sampling is a recent development that enables one to provide more structure to the collected sample items. This approach to data collection was first proposed by McIntyre (1952) for situations where taking the actual measurements for sample observations is difficult (e.g., costly, destructive, time-consuming), but mechanisms for either informally or formally ranking a set of sample units is relatively easy For discussions of some of the settings where ranked set sampling techniques have found application. The RSS procedure consists of drawing *m* random samples, each of size *m*, from the population. The *m* units of each set are ranked visually or by a negligible cost method. Then m measurements are obtained by quantifying the i^{th} visual order statistic from the i^{th} set. The whole process may be repeated *n* times until a total of nm^2 elements have been drawn from the population but only *nm* elements have been measured. The *nm* measured observations are referred to as the RSS. The procedure consists of drawing *m* random samples, each of size *m*, from the population. Takahasi and Wakimoto (1968), provided the statistical theory for RSS procedure. Assuming the sampling is from an absolutely continuous distribution, they showed that the mean of the RSS, $\hat{\mu}$, is the unbiased estimator for the population mean, μ, and has higher efficiency than the mean of SRS, $\hat{\mu}$. In order to reduce the time and the cost, the use of the RSS in order to

estimates EIV model was proposed by Al-Radaideh *et al*. (2009), they use the idea of the *L ranked set sampling* (LRSS) with Wald- type estimator to reduce the cost and increase the efficiency of the EIV model

In this paper we reintroduce the use of the RSS and the LRSS with Wald- type estimation method and with Thiel's estimation method, in order to have less error's when we estimate the EIV parameter's; moreover we made a comparison between the estimation methods itself to see which estimation would give a better estimation for the EIV model parameters in two cases, when we have outlier cases in the data, and the other without having any outliers in the data.

This paper divided into six section, the next section talk about the estimation methods that used in order to estimate the EIV model parameter, section four describe the LRSS for the bivariate data, in sections five we present the simulation study, and the conclusions is presented in the last section.

2. Grouping and Ranked Methods

Grouping methods first proposed by Wald (1940) which is simpler than the other methods and hence/or otherwise may easily be applied in many practical cases. The grouping estimators are well known as wald-type estimators. Suppose that we have two variables *x*, *y*, and *n* pairs$\{(x_i, y_i)$, $i=1,2,\ldots,n\}$, then the grouping methods can be described by the following steps:

1- Order the *n* pairs (x_i, y_i) by the magnitude of x_i; where $i=1,\ldots, n$.

2- Select proportions P_1 and P_2 such that $P_1 + P_2 \leq 1$, place the first nP_1 pairs in group one (G_1), and the last nP_2 pairs in another group (G_3), and discarding G_2 the middle group of observations; that is to say:

Place

$$(x_i, y_i)' \text{ in } \begin{cases} G_1 & if \quad x_i \leq x_{p_1} \\ G_2 & if \quad x_{p_1} < x_i \leq x_{1-p_2} \\ G_3 & if \quad x_i > x_{1-p_2} \end{cases} \qquad (4)$$

where x_{p_i} is the p_i percentile. The slope can be estimated or formulated as follows:

$$\hat{\beta} = \frac{P_2^{-1}\sum_{G_3} Y_{np_2(i)} - P_1^{-1}\sum_{G_1} Y_{np_1(i)}}{P_2^{-1}\sum_{G_3} X_{np_2(i)} - P_1^{-1}\sum_{G_1} X_{np_1(i)}}$$

, $i=1,\ldots\ldots,n$ (5)

Consequently, the intercept estimator will be:

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} \qquad (6)$$

Noting that when $P_1=P_2=1/2$ then the grouping method named by two groups (Wald ,1940), and when $P_1=P_2=1/3$ then the method is called three groups, (Nair and Shrivastava ,1942) and (Bartlett ,1949).

Theil (1950) proposed the ranked estimator, which the main idea is to find all possible slopes formulated between any two observations. Then, the median of all slopes will be the estimate. The estimates are given as follows:

$$\hat{\beta} = median\left(\hat{\beta}_{ij}\right) \quad (7)$$

Where

$$\hat{\beta}_{ij} = \frac{y_i - y_j}{x_i - x_j} \text{ , } \forall \text{i=1,...,n-1, j=2,...,n} \quad (8)$$

and

$$\hat{\alpha} = median\left(y_i - \hat{\beta} x_i\right), i = 1,...,n \quad (9)$$

3. L Ranked Set Sampling for Bivariate Data

There are varieties of techniques by which the sample may be selected. For each technique, rough estimates of the sample size can be made from knowledge of the degree of the precision desired; the related costs and time involved for each technique are also compared before making a decision (Cochran, 1997).

The simplest form of random sampling and the most used sampling methods in the practical studies is called Simple Random Sampling, in which all the units in the population have equal chance to be chosen, this sampling method does not attempt to reduce the effect of the data variation or the error estimation. McIntyre (1952) introduced a sampling method known as ranked set sampling (RSS). This technique was introduced as an efficient alternative method to simple random sampling method. Al-Nasser (2007) proposed a robust procedure based on LRSS for detecting outliers, which is a generalization for many types of ranked set sampling that introduced in the literature for estimating the population mean. Later, Al-Nasser and Al-Radaideh (2008) used LRSS to fit simple linear regression.

Assume that *(X, Y)* is a bivariate random vector such that variable *Y* is difficult to be measured, but the concomitant variable *X*, which is collected with *Y*, is easier to measure. Then one can follow the steps below in order to select LRSS:

STEP1: Randomly draw *m* independent sets each containing *m* bivariate sampling units.

STEP2: Rank the units within each sample with respect to the $X's$ by visual inspection or any other cheap method.

STEP3: Select LRSS coefficient, $k = \lfloor mp \rfloor$ such that $0 \le p < 0.5$, and $\lfloor Z \rfloor$ the largest integer value less than or equal to Z.

STEP4: For each of the first $(k+1)$ ranked samples; select the unit with rank $k+1$ and measure the y value that corresponding to $x_{(k+1)i}$ and denote it by $y_{(k+1)i}$.

STEP5: For $j = k+2,....,m-k-1$, the unit with rank j in the j^{th} ranked sample is selected and measures the y value that corresponds.

STEP6: The procedure continued until $(m-k)^{th}$ unit selected from the each of the last $(m-k)^{th}$ ranked samples, with respect to the first characteristic and measure the correspond y value.

It can be noted that, for any sample size, when $K=0$, then this steps will lead to the novel RSS, McIntyre's procedure. Also for a sample of size n with $K=\left[\frac{n-1}{2}\right]$, then the selected sample will gives a sample as in the Median Ranked Set Sampling (MRSS), adding for that the percentile Ranked Set Sampling (PRSS) can be also considered as special case of the LRSS scheme.

4. Grouping and Ranked Estimators Based on Ranked Data:

In this section we introduce the two groups, three groups, and Theil's estimates using the LRSS. In order to fit the EIV model, let:

$$L^{x}_{(i)j}=\begin{cases} X_{(k+1)j} & i\le k+1 \\ X_{(i)j} & k+2\le i\le m-k-1\ ;\ j=1,2,..,r \\ X_{(m-k)j} & m-k\le i\le m \end{cases} \tag{10}$$

be LRSS of *X* variable, and

$$L^{y}_{[i]j}=\begin{cases} Y_{[k+1]j} & i\le k+1 \\ Y_{[i]j} & k+2\le i\le m-k-1\quad ;\ j=1,2,...,r \\ Y_{[m-k]j} & m-k\le i\le m \end{cases} \tag{11}$$

Grouping estimation methods in EIV models, it classifies the observations into a finite number of groups and the centres of gravities (average) of these groups in a scatter diagram are joined together by a line to find the slope. Wald (1940) proposed grouping estimators in which he divides the observations into two equal groups after sorting them the corresponding observed values obtained from the response variable *Y*. the two groups estimates can be identified as

$$\begin{cases} \hat{\beta}_{LRSS_{2g}}=\dfrac{\bar{L}^{y}_{2}-\bar{L}^{y}_{1}}{\bar{L}^{x}_{2}-\bar{L}^{x}_{1}} \\ \hat{\alpha}_{LRSS_{2g}}=\bar{L}^{y}_{LRSS}-\hat{\beta}_{LRSS_{2g}}\bar{L}^{x}_{LRSS} \end{cases} \tag{12}$$

Similarly; Bartlett (1949) divides the observations into three equally groups and sorted it according to *x*. the estimates will not depend on the second group which will be removed in order to get a better and more efficient estimates the three groups estimates can be formulated as

$$\begin{cases} \hat{\beta}_{LRSS_{3g}}=\dfrac{\bar{L}^{y}_{3}-\bar{L}^{y}_{1}}{\bar{L}^{x}_{3}-\bar{L}^{x}_{1}} \\ \hat{\alpha}_{LRSS_{3g}}=\bar{L}^{y}_{LRSS}-\hat{\beta}_{LRSS_{3g}}\bar{L}^{x}_{LRSS} \end{cases} \tag{13}$$

And according to Theil (1950) the ranked estimator for the slope and the intercept based on the ranked are given by:

$$\hat{\beta} = median(\hat{\beta}_{ij}) \qquad (14)$$

Where

$$\hat{\beta}_{ij} = \frac{L_i^y - L_j^y}{L_i^x - L_j^x}, \forall i < j, i = 1,...,n-1,\ j = 2,...,n \qquad (15)$$

Then we have

$$\begin{cases} \hat{\beta} = median\ \hat{\beta}_{ij} \\ \hat{\alpha} = median(\mathbf{y} - \hat{\beta}\mathbf{x}) \end{cases} \qquad (16)$$

And it's easy to a proof that the proposed estimators based on the ranked data are unbiased estimators.

5. Simulation study

We made computer simulation to gain insight in the properties of the EIV model parameters estimates using the ranked data. The simulation was performed with *r = 5, 10* and *m = 3, 4, 5* and 6 for the SRS, RSS, and LRSS with coefficient $(k = 1)$ data sets. Using 10,000 replicates, estimates of the mean square errors and the relative efficiency were computed for each method. The model parameters are initialised as $\alpha = 1$, $\beta = 2$. The error terms ε_i, ξ_i and δ_i are generated from Normal distribution, the tables summarize the following simulated statistics:

1. $RE(.) = \dfrac{MSE_{SRS}(.)}{MSE_{(.)}(.)}$. (17)

2. . $MSE(\mathrm{mod}\,el) = \dfrac{1}{10000}\left[\sum_{i=1}^{10000}\left(\dfrac{\sum_{j=1}^{n}\hat{\varepsilon}_j^{\,2}}{(n-2)}\right)\right]$;where $\hat{\varepsilon}_i = y_i - \hat{y}_i$ (18)

5.1 Experiment 1

In this experiment we consider a simulated data from a normal distribution; then under the simulation assumption the relative efficiency (1) is evaluated. The samples was taken randomly firstly and then using SRS, RSS and LRSS methods in order to compare between these sampling schemes the estimation EIV model mean squared error where computed using three methods Theil , two group and three group method The results based on the proposed estimation methods are summarized in the following tables.

Here insert table 1

From the results we can see that the use of the LRSS was more efficient and more powerful than the use of the SRS in order to estimate the EIV model parameter using the grouping methods and Theil's method. Also it seems that the use of the RSS is almost has the same efficiency as the use of the SRS scheme and it's more efficient when we use the three group's method.

Also in order to compare between the estimation methods we made a comparison between the relative efficiency of the MSE for the EIV model using different estimation methods (grouping methods and Theil method). We find out that the use of the two and the three groups methods with all the sampling methods gives the same indication about the MSE of the EIV model, in other words the model MSE decreased while the sample size increase, but in the case of the use of Theil's estimation method we found that the model MSE is almost staple while the sample size is increased.

Here insert table 2

5.2 Experiment 2: Outlier case

Statistics provides a few tools for dealing with outliers in the data set. Some of these methods are only valid for small sample sizes, and none of them are overly reliable for statistical analysis. The key is to be careful if you have too many outliers in your data, it may be an indication that you should redo your experiment, or choose an alternative experimental method to collect your data.

One of the solutions for the outlier problem can be started from the beginning when you choose your sample to do the analysis, especially when you have prior information that the population may contain some outliers or extreme values may create some outlier cases, then to avoid to have such problems, one can use the LRSS to be sure that there will be any outlier case in the sample.

Now, to illustrate the performance of the proposed estimators in the presence of outliers, we carried out this experiment under the same simulation assumptions. The outliers are generated for each sampling method according to the quintiles method.

We simulate 10,000 samples for each sample size and calculate the RE for the EIV model parameter and for the regression model as well. The results for the EIV model parameters and average of the regression model are represented in the next tables.

Here insert table 3

The results give a very good indication that in the presents of the outlier in the population it will cause a problem if we use a sample taken from the population using the SRS method in order to estimate the EIV model parameter but in the other hand we have a powerful sampling scheme which is the LRSS and RSS which gives more efficient model estimation.
In other words the LRSS sampling method gives more efficient model estimation than the use of the SRS using the grouping and Theil's estimation method. Also it seems that the use of the RSS in the case of using three grouping and Theil's method gives more efficient model estimation as in the case of the use of the SRS scheme.

Here insert table 4

To compare again between the estimation methods in the case that the population have some outliers the comparison between the EIV models MSE using different estimation methods (grouping methods and Theil method) was made. we find out that the use of the two and the three groups methods with all the sampling methods gives the same indication about the EIV model MSE, in other words the model MSE increased while the sample size increase, but in the case of the use of Theil's estimation method we found that the model MSE is almost staple while the sample size is increased.

6. **Conclusions:**

One of the main aim's of statistics is to estimate the parameter of interest with less errors, less time, and cost's. In some application were the variable of interest is expensive or is difficult to measure according to this problem, the researcher will spend too much time and cost's at the same time for doing such research, lately some of the statisticians they found that the use of the LRSS and the RSS methods is appropriate sampling method to solve such these problems.

Here we used LRSS and RSS method in order to have samples to estimate the EIV model parameters using Grouping estimation methods and Theil's estimation method. After we did a simulation study we made comparison between the use of the samples which taken by using the LRSS or the RSS and with the samples taken by the SRS we find out that the use of the LRSS and RSS schemes will gives more accurate and more efficient estimators than the use of the SRS scheme..

The second comparison we made between the estimation methods, in general the simulation results indicate that all the estimation methods will gives more efficient estimators when we use the LRSS and the RSS instead of using the SRS method.

Moreover especially the LRSS method approved that it's a robust method to detect the outlier from the samples, and it will provide more precise estimates for the parameters when the data contains some outliers. Also as we saw it can be used as a generalization method for many kinds of RSS schemes.

Appendix

Table (1) EIV model relative efficiency using different sampling schemes

		Model Eff for RSS			Model Eff for LRSS		
Cycel	**Set size**	**2G**	**3G**	**THEIL**	**2G**	**3G**	**THEIL**
5	**3**	0.98	1.74	0.98	1.75	1.00	1.39
5	**4**	0.98	1.73	0.98	1.73	1.01	1.38
5	**5**	0.98	1.66	0.98	1.67	1.00	1.29
5	**6**	0.98	1.58	0.98	1.60	1.00	1.23
10	**3**	0.99	1.73	1.00	1.73	1.01	1.38
10	**4**	0.99	1.73	0.99	1.73	1.00	1.38
10	**5**	0.99	1.67	0.99	1.67	1.00	1.29
10	**6**	0.99	1.59	0.99	1.60	1.00	1.22

Table (2) EIV model mean squared using different estimation methods

		SRS			RSS			LRSS		
Cycel	**Set size**	**2G**	**3G**	**THEIL**	**2G**	**3G**	**THEIL**	**2G**	**3G**	**THEIL**
5	**3**	5.18	5.14	4.45	5.30	5.24	4.43	3.00	2.97	3.22
5	**4**	7.30	7.45	4.34	7.48	7.62	4.35	4.41	4.47	3.37
5	**5**	8.80	8.83	4.29	9.01	9.03	4.28	5.57	5.52	3.50
5	**6**	3.06	3.04	4.28	3.08	3.04	4.26	1.77	1.76	3.11
10	**3**	5.30	5.31	4.21	5.35	5.35	4.20	3.06	3.07	3.06
10	**4**	7.24	7.02	4.17	7.31	7.07	4.17	4.34	4.21	3.22
10	**5**	9.02	8.92	4.14	9.11	9.01	4.13	5.66	5.57	3.39
10	**6**	5.18	5.14	4.45	5.30	5.24	4.43	3.00	2.97	3.22

Table (3) EIV model relative efficiency using different sampling schemes (outlier case)

		Model Eff for RSS			Model Eff for LRSS		
Cycel	**Set size**	**2G**	**3G**	**THEIL**	**2G**	**3G**	**THEIL**
5	**3**	1.11	1.98	1.11	2.00	1.25	1.93
5	**4**	1.12	1.97	1.13	1.99	1.33	1.95
5	**5**	1.12	1.88	1.14	1.91	1.36	1.81
5	**6**	1.13	1.80	1.15	1.83	1.40	1.71
10	**3**	1.08	1.90	1.11	1.95	1.15	1.68
10	**4**	1.08	1.89	1.11	1.93	1.18	1.67
10	**5**	1.09	1.82	1.11	1.86	1.20	1.57
10	**6**	1.09	1.73	1.12	1.79	1.21	1.48

Table (4) EIV model mean squared using different estimation methods (outlier case)

		SRS			RSS			LRSS		
Cycel	**Set size**	**2G**	**3G**	**THEIL**	**2G**	**3G**	**THEIL**	**2G**	**3G**	**THEIL**
5	**3**	3.55	3.57	6.50	3.22	3.21	5.20	1.80	1.78	3.36
5	**4**	5.90	5.92	6.23	5.27	5.24	4.69	3.00	2.97	3.20
5	**5**	8.30	8.56	6.05	7.38	7.54	4.44	4.41	4.48	3.35
5	**6**	9.99	10.15	6.00	8.82	8.85	4.29	5.56	5.53	3.50
10	**3**	3.35	3.42	5.22	3.10	3.08	4.56	1.76	1.76	3.11
10	**4**	5.77	5.94	5.10	5.33	5.34	4.33	3.06	3.07	3.05
10	**5**	7.87	7.82	5.05	7.24	7.02	4.21	4.34	4.20	3.22
10	**6**	9.80	9.93	5.00	9.00	8.91	4.14	5.65	5.56	3.38

References

Al-Radaideh Ahmed; Amjad D. ALNasser; Enrico Ciavolino (2008).'Efficient Wald-Type Estimators for Simple Linear Measurement Error Model'. MTISD 2008 conference.

Al-Nasser, Amjad D. and Al-Radaideh, Ahmed. (2008). Estimation of Simple Linear Regression Model Using L Ranked Set Sampling. International Journal of Open Problems in Computer Science and Mathematics (IJOPCM). 1(1):18-33.

Al-Nasser, Amjad D. (2007), 'L Ranked Set Sampling: A Generalization Procedure for Robust Visual Sampling'. Communication in Statistics - Simulation and Computation. 36:33-34.

Bartlett, M. S. (1949). 'Fitting Straight Line when Both Variables are Subject to Error'. Biometrics. 5:207-212.

Cheng, C. L. and Van Ness, J. W. (1999) Statistical Regression with Measurement Error. Arnold: New York.

Fuller, W. A. (1987). Measurement Error Models. New York: John Wiley.

Kendall, M. G. and Stuart, A. (1979). The Advanced Theory of Statistics. Vol.2, Inference and Relationship. Griffin: London.

Madansky, A. (1959). The Fitting of straight lines when both variables are subject to error. Journal of the American Statistical Society, 55, 173-205.

McIntyre, G. A. (1952). 'A Method for Unbiased Selective Sampling Using Ranked Set Sampling'. Australian Journal Agriculture Research. 3:385-390.

Wald, A. (1940). 'The Fitting of Straight Lines if Both Variables are Subject to Error'. Annals of Mathematical Statistics. 11: 284-300.

Takahashi, K. and Wakimoto, K. (1968). On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering. Annals of the Institute of Statistical Mathematics. 20: 1-31.
Thiel, H. A Rank Invariant Method of Linear and Polynomial Regression Analysis, Indignations Math., 12(1950), 85-91.

Nair, R. K. and Shrivastava, P. M. (1942). On a Simple Method of Curve Fitting. Sanynkhya. 6(2): 121-132.

William G. Cochran (1977). Sampling Techniques.3rd Edition, John Wiley: New York.