

Handling Missing Data on Asymmetric Distribution

Ahmad M. H. Al-Khazaleh

Department of Mathematics, Faculty of Science
Al-albays University, Al-Mafraq-Jordan
ahmed_2005kh@yahoo.com

Abstract

The problem of imputation of missing observations emerges in many areas. Data usually contained missing observations due to many factors, such as machine failures and human error. Incomplete dataset usually causes bias due to differences between observed and unobserved data. This paper proposed Neyman allocation method to estimate asymmetric winsorizing mean for handling missing observations when the data follow the exponential distribution. Different values of the exponential distribution parameters were used to illustrate. A set of data from exponential distribution were generated to compare the performance of the proposed methods such as regression trend, average of the whole data, naive forecast and average bound of the holes and the proposed Neyman allocation method. The goodness-of-fit criterions used were the mean absolute error (MAE) and the mean squared error (MSE). It was found that the proposed method gave the best fit in the sense of having smaller error, in particular for a large percentage of missing observations.

Keywords: Neyman Allocation, Missing Data, Single imputation methods, Winsorized mean

1. INTRODUCTION

The attendance of missing observations in statistical survey data is an important issue to deal with (Little and Rubin, 1987). Besides missing observations, the time series data are possibly contaminated by outliers or are heterogeneous. Incomplete datasets may lead to results that are different from those that would have been obtained from a complete dataset (Hawthorne and Elliot, 2005).

This paper proposed a method of substituting these missing observations by using asymmetric winsorized mean. In this method, we prepared the whole data by dividing it into groups by using the stratified sampling to determine the boundaries among the strata. Stratification is the process of grouping members of the population into relatively homogeneous groups before sampling. The strata should be mutually exclusive where every element in the population must be assigned to only one stratum (Cyert, and Davidson, 1962). One of the main objectives of stratified sampling is to reduce the variance of the estimator and to get more statistical precision than with the simple random sampling (Cochran, 1977; Hanif, 2000; Hess et al., 1966; Sheikh and Ahmad, 2001). In this paper, Neyman allocation method is used to determine the boundaries of the groups.

The missing observations problem is an old one for analysis tasks. The waste of the data which can result from casewise deletion of missing values, oblige to propose alternative approaches. A problem frequently encounters in data collection is missing observations or observations may be virtually impossible to obtain, either because of time or cost constrains. In order to replace those observations, there are several different options available to the researchers (Pankratz, 1983; Patricia, 1994). Firstly, replace the missing observations with the mean of the series. Secondly; replace the missing observations with the naive forecast. Also; replace the missing observations with a simple trend forecast. Finally replace the missing observations with an average of the last two known observations that bound the missing observations.

2. NEYMAN ALLOCATION

Neyman allocation is a sample allocation method that may be used with stratified samples. We used the Neyman allocation to determine the stratum boundaries, when the population is highly skewed the Neyman allocation should be used (Cochran, 1977; Ahmad Mahir et al., 2007; Samiuddin et al., 1998). The suffix h denotes the stratum and i the unit within the stratum. When a population of N units are being stratified into L strata and the samples from each stratum are selected with simple random sampling an unbiased estimate of the population mean for the estimation variable x , is \bar{x}_{st} (st for stratified), where

$$\bar{x}_{st} = \frac{\sum_{h=1}^L N_h \bar{x}_h}{N} = \sum_{h=1}^L W_h \bar{x}_h \quad (2.1)$$

For the stratified random sampling, the variance of the estimate \bar{x}_{st} is

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^L \frac{N_h^2 S_h^2}{n_h} = \sum_{h=1}^L \frac{W_h^2 S_h^2}{n_h} \tag{2.2}$$

The variance depend on Neyman allocation is reduced to

$$V_{Ney}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L W_h S_h \right)^2 \tag{2.3}$$

Let x_0, x_L be the smallest and largest values of x in the population. The problem is to find intermediate stratum boundaries y_1, y_2, \dots, y_{L-1} and by differentiating with respect to the stratum boundaries y_h . We get the minimum variance $V_{Ney}(\bar{x}_{st})$ with respect to y_h , since y_h appears in the sum only in the terms of W_h, S_h and W_{h+1}, S_{h+1} (Cochran,1977). Hence we have the formula for finding the optimum stratum boundaries.

$$\frac{(y_h - \mu_h)^2 + S_h^2}{S_h} = \frac{(y_h - \mu_{h+1})^2 + S_{h+1}^2}{S_{h+1}} \quad \text{where } h = 1, 2, 3, \dots, L-1 \tag{2.6}$$

Suppose that the distribution of x be continuous with the density function, $f(x)$, $a < x < b$ in order to make L (groups) strata, the range of x is to be cut up at points $y_1 < y_2 < y_3 < \dots < y_{L-1}$. The relative frequency W_h , the population mean μ_h and the population variance S_h^2 at the h -th stratum are given by.

$$W_h = \int_{y_{h-1}}^{y_h} f(x) dx \tag{2.7}$$

$$\mu_h = \frac{1}{W_h} \int_{y_{h-1}}^{y_h} x f(x) dx \tag{2.8}$$

$$S_h^2 = \frac{1}{W_h} \int_{y_{h-1}}^{y_h} x^2 f(x) dx - \mu_h^2 \tag{2.9}$$

These equations are difficult to be solved, since μ_h and S_h^2 depend on y_h . We must use the iterative method to solve them by using computer program C++ (Ahmad Mahir et al. 2007). We divide the whole data into groups by putting (y_0, y_1) into group one where $y_0 = x_0$, (y_1, y_2) in group two and so on.

3. WINSORIZED MEAN

Winsorized mean is a winsorized statistical measurement of central tendency, like the mean and the median and even more similar to the truncated mean (Mingxin and

Yijun, 2009; Barnett and Lewis 1994), the winsorized mean eliminates the outliers at both ends of an ordered set of observations. Unlike the trimmed mean, the winsorized mean replaces the outliers with observed values, rather than discarding them. It involves the calculation of the mean after replacing the given parts of a probability distribution at the high and low ends with the most extreme remaining values. The winsorized mean is defined by

$$\bar{x}_w = \frac{1}{n} [(r+1)x_{r+1} + \sum_{i=r+2}^{n-s-1} x_i + (s+1)x_{n-s}] \quad (3.1)$$

which means that the winsorized mean is the average of the observations where the first r smallest values are replaced by the $(r+1)$ -th smallest value x_{r+1} , and the first s largest values are replaced by the $(s+1)$ -th largest value x_{n-s} (Barnett and Lewis 1994). When, we determine the suitable number of groups (must be more than two groups) depending on the whole data by using Equation (2.6). We obtain the boundaries of these groups at the first and the last groups. Consequently we count the number of observations inside these two groups. The number of observations in the first group is r and the number of observations in the last group is s . Subsequently, the asymmetrical winsorized mean can be calculated using Equation (3.1).

The data were generated from exponential distribution. Then, the whole data ordered from the smallest to the largest, the smallest value is zero and the largest value is 17.12. The data were assumed having 5%, 10%, and 20% randomly missing observations in this study.

4 THE BOUNDARIES OF GROUPS IN EXPONENTIAL DISTRIBUTION

Suppose that the time series x_1, x_2, \dots, x_n of n observations come from an exponential distribution with probability density function

$$f(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, \quad x > 0, \quad (4.1)$$

where λ is the parameter of exponential function. In order to make L (groups) strata, the domain of $f(x)$ was truncated from, $(0, \infty)$ to $(0, b]$ where b is the largest value of x . From equations (2.7), (2.8) and (2.9), the relative frequency W_h , mean μ_h and variance S_h^2 of the h -th stratum can be computed for the exponential distribution as the following

$$W_h = e^{-\lambda y_{h-1}} - e^{-\lambda y_h} \quad (4.2)$$

$$\mu_h = \frac{(y_{h-1} + \frac{1}{\lambda})e^{-\lambda x_{h-1}} - (y_h + \frac{1}{\lambda})e^{-\lambda x_h}}{e^{-\lambda y_{h-1}} - e^{-\lambda y_h}} \tag{4.3}$$

$$S_h^2 = \frac{(y_{h-1}^2 + \frac{2}{\lambda}y_{h-1} + \frac{2}{\lambda^2})e^{-\lambda x_{h-1}} - (y_h^2 + \frac{2}{\lambda}y_h + \frac{2}{\lambda^2})e^{-\lambda x_h}}{e^{-\lambda y_{h-1}} - e^{-\lambda y_h}} - (\mu_h)^2 \tag{4.4}$$

Table 1 Stratum boundaries for groups size of three to eight in the case of exponential distribution when $\lambda = 0.2$,

Boundaries	3G	4G	5G	6G	7G	8G
y ₁	3.1952	2.3026	1.8005	1.4783	1.2539	1.0887
y ₂	7.9365	5.3023	3.9989	3.214	2.6881	2.3107
y ₃		9.6199	6.8234	5.3165	4.3633	3.7033
y ₄			10.7832	7.9841	6.37757	5.3219
y ₅				11.6395	8.9041	7.2547
y ₆					12.2981	9.6541
y ₇						12.8213

Table 2 Stratum boundaries for groups size of three to eight in the case of exponential distribution when $\lambda = 0.4$,

Boundaries	3G	4G	5G	6G	7G	8G
y ₁	1.8824	1.3550	1.0594	0.8699	0.7379	0.6407
y ₂	4.9603	3.2233	2.4064	1.9241	1.6041	1.3759
y ₃		6.2623	4.2593	3.2626	2.6529	2.2383
y ₄			7.2567	5.0994	3.9826	3.2814
y ₅				8.0536	5.8028	4.6022
y ₆					8.7130	6.4054
y ₇						9.2714

Table 3 Stratum boundaries for groups size of three to eight in the case of exponential distribution when $\lambda = 0.6$

Boundaries	3G	4G	5G	6G	7G	8G
y ₁	1.2723	0.9169	0.7176	0.5898	0.5007	0.4351
y ₂	3.3727	2.1886	1.6342	1.30712	1.0903	0.9356
y ₃		4.2872	2.9051	2.2233	1.8074	1.5251
y ₄			5.0018	3.4935	2.7231	2.2419
y ₅				5.5878	3.9923	3.1570
y ₆					6.0842	4.4253
y ₇						6.5143

The number of groups or strata depends on the size of the population. If we want to divide the whole data into three groups, we need to calculate two boundaries y_1 and y_2 since $x_0 = y_0 = 0$ and $y_3 = 17.12$ which is the largest value. The boundary y_1 depend on the μ_1 and μ_2 and S_1^2 and S_2^2 , also y_2 depends on μ_2 and μ_3 and S_2^2 and S_3^2 . We calculated these values by applying equation (2.6) depending on equations (2.8) and (2.9) and in the calculation of the boundaries we assumed the parameter of exponential distribution are $\lambda = 0.2, 0.4, \text{ and } 0.6$. If we divide the whole data into four groups, we need to calculate three boundaries y_1 , y_2 and y_3 . Similarly we find the boundaries of the strata (group) for the five, six, seven and eight groups in the same way by using the C++ program. Tables 1, 2, and 4 have shown the boundaries for all groups for the different values of the parameter of exponential distribution for $\lambda = 0.2, 0.4, \text{ and } 0.6$ respectively.

5 SUMMARY AND DISCUSSION

The method to determine the two parameters of asymmetric winsorized mean (r, s) when the whole data was divided into three to eight groups as follow. We assumed that the data with different percentage of missing observations such as 5%, 10%, and 20% which has been selected randomly. In the first case when the data has 5% missing observations, the asymmetric winsorized mean was calculated as the following: Suppose the stratum boundaries for the whole data was divided into three groups with 5% missing data with $\lambda = 0.2$. The first group has the interval $(0, 3.1952)$, the second group has the interval $(3.1952, 7.9365)$ and the third group has the interval $(7.9365, 17.12)$ (see Table 1). Since we are interested in the first and the last group, then we count the number of observations in both groups. The number of observations in the first one was 119, and the number of observations in third group was 55 as shown in Table 5a. Finally the asymmetric winsorized mean was calculated by taking $r = 119$ and $s = 55$ and by applying equation (3.1), which is equal to 5.0037. The same procedure to compute the asymmetric winsorized mean was followed for different groups for 5% missing observations as shown in tables 5b, and 5c.

Table 5a The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.2$, and 5% missing observations for three groups to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	119	65	46	32	25	21
s	55	25	13	11	9	8
$W(r,s)$	5.0037	4.9717	4.9573	4.9303	4.9418	4.9492

Moreover, for the data with 10% missing observations, when the whole data was divided into three groups with $\lambda = 0.2$, the number of observations in the first group was 119 which was the value of r and the number of observations in the last one 50 which was s , then the asymmetric winsorized mean was equal to 4.9483. In the same way, the asymmetric winsorized mean was calculated for different groups. Tables 6a, 6b, and 6d have shown the numbers of observations in the first and last groups for different values of exponential distributions parameter and the values of asymmetric winsorized mean.

Table 5b The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.4$ and 5% missing observations for three groups to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	48	28	21	17	13	12
s	170	109	76	52	42	32
$W(r,s)$	3.9331	4.3423	4.5547	4.6725	4.7491	4.8023

Table 5c The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.6$ and 5% missing observations for three groups to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	26	18	12	12	11	9
s	267	208	169	139	116	100
$W(r,s)$	2.9232	3.5130	3.8626	4.0815	4.2451	4.3604

Table 6a The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.2$, and 10% missing observations for three groups to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	119	67	48	31	24	20
s	50	22	12	10	8	7
$W(r,s)$	4.9483	4.8934	4.8633	4.8391	4.8491	4.8555

Table 6b The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.4$ and 10% missing observations for three groups to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	49	27	20	16	13	12
s	156	100	69	47	38	29
$W(r,s)$	3.8825	4.2721	4.4784	4.5914	4.6656	4.7153

Table 6c The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.6$ and 10% missing observations for three groups to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	25	17	12	11	10	9
s	246	192	155	128	106	92
$W(r,s)$	2.9366	3.4768	3.8089	4.0170	4.1728	4.2903

Finally, for 20% missing observations, the data was divided into three groups. The number of the observations of the first group was 103 which was the value of the parameter r and the number of observations in the third group was 51 which was the value of parameter s , then the asymmetric winsorized mean computed was found to be 5.0346, as shown in Table 7a. Similarly the asymmetric winsorized mean computed for the other groups with 20% percentage of missing observations as obtained in Tables 7b, and 7c.

Table 7a The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.2$, and 20% missing observations for three groups up to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	103	59	42	29	23	19
s	51	25	13	11	9	8
$W(r,s)$	5.0346	5.0074	4.9975	4.9692	4.9892	4.9948

Table 7b The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.4$ and 20% missing observations for three groups up to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	43	25	19	15	12	11
s	145	96	68	49	40	32
$W(r,s)$	3.9075	4.3226	4.5480	4.6763	4.7635	4.8225

Table 7c The values of the two parameters and the asymmetric winsorizing mean when $\lambda = 0.6$ and 20% missing observations for three groups up to eight groups

Parameter	3G	4G	5G	6G	7G	8G
r	23	17	11	10	9	8
s	241	174	144	121	101	89
$W(r,s)$	2.9646	3.4945	3.8324	4.0508	4.2227	4.3425

We compared the result of using asymmetric winsorized mean method for missing observations with other estimation methods for missing observations. The first method is to replace the missing observations with the mean of the series. This mean can be calculated over the entire range of the sample. Second method is to replace the missing observations with the naive forecast. Naive model is the simplest form of a univariate forecast model. This model uses the current time period's value for the next time period, that is $\hat{X}_{t+1} = X_t$. Also, we can replace the missing observations with a simple trend forecast. This is accomplished by estimating the regression equation of the form $X_t = a + bt$ (where t is the time) for the periods prior to the missing value. Then use the equation to fit the time periods missing. Finally, we can replace the missing observations with an average of the last two known observations that cover the missing observations (Patricia,1994).

The accuracy of estimating the missing observations with asymmetric winsorized mean depends on how close the estimating values to the actual values. In practice, we define the difference between the actual and the estimating values as an error. If the estimation is doing a good job, the error will be relatively small. This means that the error for each time period is purely random fluctuation around original value. So we should get a value equal 0 or near 0. We tested asymmetric winsorized mean approach to estimating missing data points on time series data sets with respect to the other methods. In this paper the following statistical measures were used as the estimators of accuracy (Patricia, 1994).

1- The mean absolute errors

$$MAE = \frac{\sum_{t=1}^n |e_t|}{n} \tag{5.1}$$

2- The mean square errors

$$MSE = \frac{\sum_{t=1}^n e_t^2}{n} \tag{5.2}$$

where n is the number of imputations, e_i is difference between the actual and the estimating values. In order to evaluate the accuracy of the estimation procedure for missing observations, we used mean absolute error (MAE) as in Equation (5.1) and mean square error (MSE) as in Equation (5.2). Tables 8a to 9c shown the amount of the MAE and MSE for the estimation of the missing observations by using the proposed method, and Tables 11 and 12 for the other methods. The MAE values as shown in Tables 8a to 8c for the proposed methods and Table 10 for the other methods. At the 5% missing observations case, the simple average method does better than the asymmetric winsorized methods when $\lambda = 0.2$, in three and four groups, but for the other values, it is clear that the proposed method is better than the other methods. For the 10% missing observations, the average coverage method does better than the other methods, while the simple average method does better than the asymmetric method only when $\lambda = 0.2$, at three and four groups, $\lambda = 0.4$ at three groups and when $\lambda = 0.6$ at three, four and five groups. Finally, for the 20% missing observations, there is only one case; the simple average method does better than the asymmetric winsorized method when $\lambda = 0.2$, in three groups.

For the second accuracy measures, the MSE as shown in Tables 9a to 9c for the proposed methods and Table 11 for the other methods. For the 5% missing observations case in general, the asymmetric winsorized mean methods do better than the others, but the simple average method do better than the proposed method when $\lambda = 0.2$, in three and four groups. For the 10% missing observations, the asymmetric winsorized methods do better than the naive and trend methods, while the average bound method do better than the proposed method, the mixed results between the proposed methods and the simple average method. Finally, for the 20% missing observations, the simple average method does better than the proposed method in two cases when the $\lambda = 0.2$, in three groups and $\lambda = 0.6$ in three and four groups.

In this paper, 18 asymmetric winsorized mean methods were proposed to handle missing observations with 5%, 10%, and 20% missing observations. General results shown that the proposed methods do better than the other methods, in particular when the data having 20% missing observations. We recommend using the asymmetric winsorized method when the data have more missing observations and dividing the whole data into more than four groups.

Table 8a The values of mean absolute error of asymmetric winsorized mean for 5% missing data, $\lambda = 0.2, 0.4, \text{ and } 0.6$ for different groups

	MAE	3G	4G	5G	6G	7G	8G
$\lambda = 0.2,$		2.9554	2.9403	2.9335	2.9207	2.9261	2.9296
$\lambda = 0.4$		2.4483	2.6421	2.7428	2.7986	2.8349	2.8600
$\lambda = 0.6$		2.1037	2.2804	2.4149	2.5186	2.5961	2.6507

Table 8b The values of mean absolute error of asymmetric winsorized mean for 10% missing data, $\lambda = 0.2, 0.4,$ and 0.6 for different groups

	MAE	3G	4G	5G	6G	7G	8G
$\lambda = 0.2,$		2.4763	2.4711	2.4682	2.4659	2.4669	2.4675
$\lambda = 0.4$		2.5027	2.4623	2.4525	2.4522	2.4546	2.4569
$\lambda = 0.6$		2.7818	2.5967	2.5167	2.4831	2.4679	2.4615

Table 8c The values of mean absolute error of asymmetric winsorized mean for 20% missing data, $\lambda = 0.2, 0.4,$ and 0.6 for different groups

	MAE	3G	4G	5G	6G	7G	8G
$\lambda = 0.2,$		1.9317	1.9249	1.9224	1.9154	1.9204	1.9218
$\lambda = 0.4$		1.8533	1.8311	1.8396	1.8559	1.8711	1.8819
$\lambda = 0.6$		2.0876	1.9343	1.8652	1.8384	1.8311	1.8311

Table 9a The values of mean square error for asymmetric winsorized mean for 5% missing data, $\lambda = 0.2, 0.4,$ and 0.6 for different groups

	MSE	3G	4G	5G	6G	7G	8G
$\lambda = 0.2,$		10.575	10.455	10.401	10.302	10.344	10.371
$\lambda = 0.4$		7.6499	8.4974	9.0695	9.4255	9.6721	9.8500
$\lambda = 0.6$		6.9912	7.1339	7.5377	7.9186	8.2658	8.5426

Table 9b The values of mean square error for asymmetric winsorized mean for 10% missing data, $\lambda = 0.2, 0.4,$ and 0.6 for different groups

	MSE	3G	4G	5G	6G	7G	8G
$\lambda = 0.2,$		9.3746	9.3545	9.3460	9.3405	9.3426	9.3441
$\lambda = 0.4$		10.061	9.5468	9.3974	9.3515	9.3354	9.3307
$\lambda = 0.6$		12.5735	10.9198	10.1925	9.8494	9.6491	9.5303

Table 9c The values of mean square error for asymmetric winsorized mean for 20% missing data, $\lambda = 0.2, 0.4,$ and 0.6 for different groups

	MSE	3G	4G	5G	6G	7G	8G
$\lambda = 0.2,$		5.5162	5.4781	5.4645	5.4270	5.4534	5.4609
$\lambda = 0.4$		5.1749	5.0050	5.0572	5.1322	5.2020	5.2578
$\lambda = 0.6$		6.8411	5.6858	5.2424	5.0773	5.0144	5.0055

Table 10 The values of mean absolute error for the other methods for 5, 10 and 20%.

missing	MAE	Simple Average	Naive	Bound Average	Trend Regression
5%		2.93538	4.033724	3.196569	3.432913
10%		2.46873	3.231142	2.195116	2.985184
20%		1.92651	2.520092	2.366819	4.033494

Table 11 The values of mean square error for the other methods for 5, 10 and 20%.

missing	MSE	Simple Average	Naive	Bound Average	Trend Regression
5%		10.41615	24.90699	15.05952	15.41698
10%		9.347366	16.18883	8.393193	15.04977
20%		5.486735	10.62327	9.037119	28.30867

REFERENCES

- [1] A.K. Sheikh, and M. Ahmad,. Statistical Models of Accelerated Life Testing. Pak. J. Statist. 17 (2): (2001) 75-102.
- [2] A. Pankratz,. Forecasting with Univariate Box - Jenkins Models. John Wiley, USA, 1983.
- [3] E. G. Patricia, Introduction to Time Series Modeling and Forecasting in Business and Economics. McGraw-Hill, Inc. New York, 1994.
- [4] G. Hawthorne, and P. Elliot,. Imputing Cross-Sectional Missing Data: Comparison of Common Techniques. Australian and New Zealand Journal of Psychiatry. 39: (2005) 583-590.
- [5] I. Hess, V.K. Sethi, and T.R. Balakrishnan, Stratification: a practical investigation. J. Amer. Statist. Assoc. 61, (1966) 74-90.
- [6] M. Hanif,. Design and Model-based sampling inference. Pak. J. Statist. 16(3): (2000) 229-246.
- [7] M. Samiuddin, M. Hanif, and A.K.A Kattan, An optimum form for the design based ratio estimator. Pak. J. Statist. 14 (1) (1998) 81-96.
- [8] R. Ahmad Mahir, A.M.H. Al-Khazaleh and M.M.T. Al-Kassab,. Approximation Method in Finding Optimum Stratum Depending on Neyman Allocation Applied on Beta Distribution. Proceedings of the 12th WSEAS International Conference on Applied Mathematics. Cairo, Egypt. WSEAS Press. (2007) 341-345.

- [9] R.M. Cyert, and H.J. Davidson, Statistical Sampling for Accounting Information. Prentice-Hall, Englewood Cliffs, NJ, pp. (1962) 116–127,.
- [10] R. J. A. Little, and B.B. Rubin,. Statistical Analysis with Missing Data. John Wiley: New York, 1987.
- [11] V. Barnett, and T. Lewis Outliers in Statistical Data. 3rd Edition, John Wiley, New York, USA,1994.
- [12] W.G. Cochran. Sampling Techniques. 3rd Edition. John Wiley, New York, USA, 1977.
- [13] W. Mingxin, and Z. Yijun, . Trimmed and winsorised means based on a scaled deviation. J. of Stat. Planning and Inference. 139: (2009) 350 – 365.

Received: September, 2012