Evaluation of Accuracy of the Estimation Methods for Replacing Missing Values for Time Series Variables. Using the statistical packages software (SPSS & MINITAB)

Dr. Marie M. Mahmoud
Dept. of Statistics & Mathematics
Faculty of Commerce, Tanta University

ABSTRACT

Missing values define specified data values as user-missing. It is often useful to know why information is missing. For example, you might want to distinguish between data missing because a respondent refused to answer and data missing because the question didn't apply to that respondent. Data values specified as user-missing are flagged for special treatment and are excluded from most calculations. You can enter up to three discrete (individual) missing values, a range of missing values, or a range plus one discrete value. Ranges can be specified only for numeric variables. In this paper, we aim at to determine the estimation methods using SPSS and MINITAB packages for replacing missing values for time series variables, which minimize the total prediction error (direct forecasting error and indirect forecasting error or replacing processing error), according to three Measures of Accuracy of the fitted model: 1) can Absolute Percentage Error (MAPE), Mean Absolute Deviate, a (MAD), and Mean Squared Deviation (MSD) in Time Series. In this paper, we consider the difference between the continuous and discrete distributions, also the difference between the symmetric and asymmetric distributions, and so the difference between the moderate and heavy tailed distributions.

Keywords: Time Series Data, Missing Values, Series mean, Mean & Median of nearby points, linear interpolation, linear trend at point, Measures of Accuracy, and Forecasting Error.
(1). INTRODUCTION
Reducing the estimator bias is very important in statistical analysis. In this paper, five methods for replacing missing values for time series variables are applied, and we aim at to determine the best estimation method, according to: 1- The probability distribution of the variable:
2- Three measures of accuracy of the fitted model.

To set the scene, I first need to briefly describe what I mean, by:

**Time Series Data:** Corresponds to the sequence of values for a single variable in ordinary data analysis. Each case (row) in the data represents an observation at a different time. The observations must be taken at equally spaced time intervals.

**Creating Time Series:** Creating Time Series means creates new variables bated on functions of existing numeric time series variables. These transformed values are useful in many time series analysis procedures.

**User-Mining Values:** Values you have specified as missing, using Define Variable in the Data menu. You can specify individual missing values for numeric or string variables or a range of missing values for numeric variables. See also system-missing values.

**System-Missing Values:** Values assigned by the program when values in your data are undefined according to the format type you have specified, when a numeric field is blank, or when a value resulting from a transformation command is undefined. Numeric system-missing values are displayed as periods. String variables cannot have system-missing values, since any character is legal in a string variable.

**Forecasting Error:** The discrepancy between an observed value and its forecast, based on a specified model.

**In SPSS:** To Create a New Time Series Variable

**From the menus choose:** 1- Transform 2- Create Time Series...

- Select the time series function you want to use to transform the original variable(s).
- Select the variable(s) from which you want to create new time series variables. Only numeric variables can be used.

**Optionally, you can:**
- Enter variable names to override the default new variable names.
- Change the function for a selected variable.

**In SPSS:** Default new variable names are the first six characters of the existing variable used to create it, followed by an underscore and a sequential number. For example, for the variable ECONOMETRIC, the new variable name would be ECONOM_1.

**In SPSS:** To Define Missin e Values for a Variable

- Make the Data Editor the active window.
- If the Data view is displayed, double-click the variable name at the top of the column in the Data view or click the Variable View tab.
- Click the button in the Missing cell for the variable you want to define.
- Enter the values or range of values that represent missing data.
**Missing values for string variables:** All string values, including null or blank values, are considered valid values unless you explicitly define them as missing. To define null or blank values as missing for a string variable, enter a single space in one of the fields for Discrete missing values. You cannot define missing values for long string variables (string variables longer than eight characters).

**In MINITAB:** Missing value symbol: Minitab uses an asterisk (*) in (numeric columns) and a blank in (alpha columns) to represent missing values in a column. When you enter * as a value in the Data window or at the DATA> prompt in the Session window, you do not have to enclose it in quotation marks. However, you must enclose the * in quotation marks when it is part of the command line. Most commands exclude from analysis all rows with a missing value and display the number of excluded points. When an arithmetic command operates on a missing value, Minitab sets the result to *.

**Sample from Columns**
Randomly samples rows from one or more columns. You can sample with replacement (the same row can be selected more than once), or without replacement (the same row is not selected more than once).

**Dialog box items**
- **Sample rows from column(s):** Specify the number of rows to randomly select, and then select the column(s) you want the data to be sampled from.
- **Store samples in:** Specify the column(s) where the sampled values are to be stored. If you sample from several columns at once, they must all have the same length; Minitab selects the same rows from each column.
- **Sample with replacement:** Check to sample with replacement. If left unchecked, Minitab samples without replacement, so the same row cannot be selected more than once.

**To select a Random Sample from columns.**
You can randomly sample rows from one or more columns.
1. Choose Calc > Random Data > Sample from Columns.
2. In Sample, enter the number of rows you want to sample.
3. In the box following rows from column(s), enter equal-length Columns from which to sample.
4. In Store samples in, enter the columns in which you want to store the sample data. Click OK.

**Text data:** Minitab handles numeric data, text data (formerly called aloha data), and date/time data.

**Rules for text data include:**
1. Text data in a column may be up to 80 characters long.
2. You may use any characters (letters, numbers, punctuation symbols, and blanks).
3. No column can contain both text and numeric data:
   - Minitab treats numbers appearing in any text column as text characters.
   - Some Minitab commands do not handle text data. You can convert text to other types of data.
(2) ESTIMATION METHODS for REPLACING MISSING VALUES

Suppose that the values of a time series are:
X₁, X₂, X₃, ......................... X₀

Missing observations can be problematic in analysis, and some time series measures cannot be computed if there are missing values in the series. Replace Missing Values creates new time series variables from existing ones, replacing missing values with estimates computed using one of these methods.

(2.1) - Series mean.
Replacing missing values with the mean for the entire series. All missing values are replaced by (K) where:
K = The mean of time series ΣX / n, 2, ...., n

(2.2) - Mean of nearby points.
Replacing missing values with the mean of valid surrounding values. The span of nearby points is the number of valid values above and below the missing value used to compute the mean.

Let Xᵢ be missing value. Then X₁ is replaced by (K) where:
K = (X₁₋₁ + Xᵢ₋₁ + Xᵢ₋₂ + Xᵢ₋₃ + Xᵢ₋₄) / 4  (If the span of nearby points = 2), and 2 < i < n-2;
K = (X₁₋₃ + X₁₋₁ + X₁+1 + X₁+2 + X₁+3) / 8  (If the span of nearby points = 4), and 4 < i < n-4;
K = median of (X₁₋₄, X₁₋₁, X₁₋₂, X₁₋₃, Xᵢ₋₃, Xᵢ₋₂, Xᵢ₋₁, X₁+1, X₁+2, X₁+3)  (If the span of nearby points = g), and g < i < n-g

(2.3) - Median of nearby points.
Replacing missing values with the median of valid surrounding values. The span of nearby points is the number of valid values above and below the missing value used to compute the median.

Let Xᵢ be the missing value. Then Xᵢ is replaced by (K) where:
K = median of (X₁₋₂, X₁₋₁, X₁₋₀)  (If the span of nearby points = 2), and 2 < i < n-2;
K = median of (X₁₋₂, X₁₋₁, X₁₋₀, X₁+1, X₁+2, X₁+3, X₁+4)  (If the span of nearby points = 4), and 4 < i < n-4;
K = median of (X₁₋₅, X₁₋₄, X₁₋₃, X₁₋₂, X₁₋₁, Xᵢ₋₀)  (If the span of nearby points = g), and g < i < n-g
(2-4) - Linear interpolation. Replacing missing values using a linear interpolation. The last valid value before the missing value and the first valid value after the missing value are used for the interpolation. If the first or last case in the series has a missing value, the missing value is not replaced.

Let \( X_i \) be a missing value. Then \( X_i \) is replaced by \( K \) where:

\[
K = X_{i-1} + \frac{x_{i-1}}{2} x_{i+1} - \frac{x_{i-1}}{2}
\]

\[ \text{which } l < i \]

Then, \( K = \frac{x_{i-1}}{2} + \frac{x_{i+1}}{2} \) which \( l < i \)

But note that:

a - Linear interpolation = Mean of nearby points (If the span of nearby points = 1, and there is no sequential values are missing).

b - If \( X_i \), \( X_{i+1} \) are missing values. Then \( X_i, X_{i+1} \) are replaced by

\[
K_i = X_{i+1} + \frac{X_{i+1}}{2} - \frac{X_{i-1}}{2}
\]

\[ = \frac{X_{i+1} + X_{i-1}}{2} - \frac{X_{i-1}}{2}
\]

\[ = \frac{X_{i+1} - X_{i-1}}{2} \]

(2-5) - Linear trend at point. Replacing missing values with the linear trend for that point. The existing series is regressed on an index variable scaled 1 to n. Missing values are replaced with their predicted values.

The steps of Replacing Missing Values:

I - Suppose that the observations of a time series (missing and nonmissing values) are:

\( X_1, X_2, X_3, \ldots, X_0-2, X_{p-1}, X_{p} \)

And consider \( X \) is a dependent (response) variable.

II - Let \( Y \) be an independent (predictor) variable, and \( Y \)'s values are: \( 1, 2, 3, \ldots \)
III — after excluding the rows of missing values from X, Y columns,

Determine the regression equation: $X = a + b \cdot Y$

Let $X_i$ be missing value. Then $X_i$ is replaced by $(K_i)$ where:

\[ K_i = a + b \cdot i \]

...... (20)

**Missing Values in Functions:** Functions and simple arithmetic expressions treat missing values in different ways.

**In the expression:** $(\text{var}1+\text{var}2+\text{var}3)/3$

The result is missing if a case has a missing value for at least one of the three variables.

**In the expression:** $\text{MEAN} (\text{var}1, \text{var}2, \text{var}3)$

The result is missing only if the case has missing values for all three variables.

For statistical functions, you can specify the minimum number of arguments that must have nonmissing values. To do so, type a period and the minimum number after the function name, as in $\text{MEAN.2} (\text{var}1, \text{var}2, \text{var}3)$

**In SPSS:** To Replace Missing Values for Time Series Variables

- From the menus choose:
  - Transform
  - Replace Missing Values...
    - Select the estimation method you want to use to replace missing values.
    - Select the variable(s) for which you want to replace missing values.
  - Optionally, you can:
    - Enter variable names to override the default new variable names.
    - Change the estimation method for a selected variable.

---

**(3) TIME SERIES TRANSFORMATION FUNCTIONS**

In this section we will illustrate the meaning of the system-missing value for the variable in many cases of transformations:

**I - Difference.** Non seasonal difference between successive values in the series. The order is the number of previous values used to calculate the difference. Because one observation is lost for each order of difference, system-missing values appear at the beginning of the series. For example, if the difference order is 2, the first two cases will have the system-missing value for the new variable.
2 - **Seasonal difference.** Difference between series values a constant span apart. The span is based on the currently defined periodicity. To compute seasonal differences, you must have defined date variables (Data menu, Define Dates) that include a periodic component (such as months of the year). The order is the number of seasonal periods used to compute the difference. The number of cases with the system-missing value at the beginning of the series is equal to the periodicity multiplied by the order. For example, if the current periodicity is 12 and the order is 2, the first 24 Cases will have the system-missing value for the new variable.

3 - **Centered moving average.** Average of a span of series values surrounding and including the current value. The span is the number of series values used to compute the average. If the span is even, the moving average is computed by averaging each pair of uncentered means. The number of cases with the system-missing value at the beginning and at the end of the series for a span of n is equal to n/2 for even span values and for odd span values. For example, if the span is 5, the number of cases with the system-missing value at the beginning and at the end of the series is 2.

4 - **Prior moving average.** Average of the span of series values preceding the current value. The span is the number of preceding series values used to compute the average. The number of cases with the system-missing value at the beginning of the series is equal to the span value.

5 - **Running median.** Median of a span of series values surrounding and including the current value. The span is the number of series values used to compute the median. If the span is even, the median is computed by averaging each pair of uncentered medians. The number of cases with the system-missing value at the beginning and at the end of the series for a span of n is equal to n/2 for even span values and for odd span values. For example, if the span is 5, the number of cases with the system-missing value at the beginning and at the end of the series is 2.

6 - **Cumulative sum.** Cumulative sum of series values up to and including the current value.

7 - **Lag.** Value of a previous case, based on the specified lag order. The order is the number of cases prior to the current case from which the value is obtained. The number of cases with the system-missing value at the beginning of the series is equal to the order value.
8 - Lead. Value of a subsequent case, based on the specified lead order. The order is the number of cases after the current case from which the value is obtained. The number of cases with the system-missing value at the end of the series is equal to the order value.

9 - Smoothing. New series values based on a compound data smoother. The smoother starts with a running median of 4, which is centered by a running median of 2. It then resmoothes these values by applying a running median of 5, a running median of 3, and harming (running weighted averages). Residuals are computed by subtracting the smoothed series from the original series. This whole process is then repeated on the computed residuals. Finally, the smoothed residuals are computed by subtracting the smoothed values obtained the first time through the process. This is sometimes referred to as T4253H.smoothing.

(4) MEASURES of ACCURACY

Accuracy refers to the closeness of the measurements to the "actual" or "real" value of the physical quantity. Therefore, an "accurate" estimate has small bias.

Minitab computes three measures of accuracy of the fitted model: \( \text{MAPE, MAD, and MSD} \) for each of the simple forecasting and smoothing methods. For all three measures, the smaller the value, the better the fit of the model. Use these statistics to compare the fits of the different methods.

(I) Mean Absolute Percentage Error (MAPE)

\[
\text{MAPE} = \left( \frac{\sum |Y_t - \hat{Y}_t|}{\sum Y_t} \right) + n \times 100
\]

Where: \( Y_t \) equals the actual value; \( \hat{Y}_t \) equals the forecast value; \( n \) equals the number of forecasts; \( t = 1, 2, \ldots, n; \) and \( Y_t \neq 0 \)

(MAD) Mean Absolute Deviation (MAD)

\[
\text{MAD} = \sum |Y_t - \hat{Y}_t| + n
\]
Mean Squared Deviation (MSD) in Time Series

MSD is very similar to MSE, mean squared error, a commonly-used measure of accuracy of fitted time series values. Because MSD is always computed using the same denominator, \( n \), regardless of the model, you can compare MSD values across models. Because MSE are computed with different degrees of freedom for different models, you cannot always compare MSE values across models.

\[
MSD = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(23)

(5) – NUMERICAL STUDY

Methods dealing with analysis of data with missing values (Incomplete Data) can be classified into:
- Analysis of complete cases, including weighting adjustments,
- Imputation methods, and extensions to multiple imputation, and
- Methods that analyze the incomplete data directly without requiring a rectangular data set, such as maximum likelihood and Bayesian methods.

In this section:
- We conduct simulations designed to investigate what the effect on the forecasting accuracy of the time series model, when using various Estimation Methods for Replacing Missing Values, various probability distributions, and various sample sizes.
- The simulations were carried out using two statistical packages (MINITAB and SPSS). The random data generator used in the simulations is that from the MINITAB calc Toolbox, and the Estimation Methods for Replacing Missing Values used in the numerical study is that from the SPSS Transform Toolbox.
- We choose three continuous Distributions (2 symmetric distributions and one assymmetric distribution), and one Discrete Distribution for our simulation study.
- Our symmetric distributions are the standard normal distribution \( N(0, 1) \), and the Student \( t \) — distribution with 5 degrees of freedom \( t(5) \). These represent moderate and very heavy tailed distributions.
- For the assymmetric distribution we use the log-normal distribution, with \( \mu = 0 \), \( \sigma = 1 \), denoted \( \text{Log-N}(0, 1) \). (A variable \( X \) has a lognormal distribution if \( \log(X) \) has a normal distribution), and For the Discrete Distribution we use the Binomial Distribution. (The probability of success = 0.5),
• For a range of values for n (50, 100, 500, and 1000), we use the four distributions and calculate the results of the various Estimation Methods for Replacing Missing Values discussed in the section-2, and the various Measures of Accuracy discussed in the previous section.

• For the Mean of nearby points and Median of nearby points Estimation Methods for Replacing Missing Values we choose the span of nearby points = 4.

• For all cases, always we consider that the number of the missing values equals to twenty percent of sample size, then (m) = (20%*n)

The steps of numerical study:

1 - Using MINITAB, to generate the random data according to the probability distribution and the sample size (n).

2- Select a Random Sample (m) from the column of the sample size in the step one, and do it as missing values, where (m) = (20%*n).

3- Using SPSS, to estimate the Missing Values by using the previous Estimation Methods.

4- Compute the measures of accuracy using:
   a - MINITAB Toolbox — 0 cafe-0 calculator, or
   b - SPSS Toolbox — Transform — compute

5 — Repeat the steps 2, 3, and 4 for another random sample from the same column (the same probability distribution and the same sample size).

6 - Compute the average of the computed measures of accuracy for all selected samples. In this paper, we select 100 Random Samples in each case.

7 - Repeat the steps 2, 3, 4, 5, and 6 for all cases.

• Our simulation results are reported in the following Tables: (1, 2, 3, and 4).

---

If N = population size, and n = sample size. Then we can select number of simple random samples (without replacement) = \( N \binom{n}{N} \), or 
(N) Combination (n).
Table (1): n = 50

<table>
<thead>
<tr>
<th>Probability distribution</th>
<th>Measures of Accuracy</th>
<th>Series mean</th>
<th>Mean of Median nearby of points(4)</th>
<th>Linear interpolation trend at point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>0.9902</td>
<td>2.3228</td>
<td>2.4054</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>0.6782</td>
<td>0.8014</td>
<td>1.1395</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.3772</td>
<td>0.3999</td>
<td>1.9232</td>
</tr>
<tr>
<td>NO, 1)</td>
<td><strong>T-5</strong></td>
<td><strong>0.3015</strong></td>
<td><strong>0.1108</strong></td>
<td><strong>0.9422</strong></td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
<td>9.5667</td>
<td>10.4932</td>
<td>15.0341</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>1.5616</td>
<td>1.5523</td>
<td>1.9950</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>1.6258</td>
<td>1.0021</td>
<td>1.0982</td>
</tr>
<tr>
<td>Log-N (0, 1)</td>
<td><strong>MAPE 0.0987</strong></td>
<td><strong>0.1899</strong></td>
<td><strong>0.1289</strong></td>
<td><strong>0.3219</strong></td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>3.4297</td>
<td>5.0143</td>
<td>52995</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>6.9329</td>
<td>11.2396</td>
<td>5.2062</td>
</tr>
<tr>
<td>Binomial</td>
<td><strong>MAPE 0.4540</strong></td>
<td><strong>0.4934</strong></td>
<td><strong>0.4331</strong></td>
<td><strong>0.3912</strong></td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>1.0057</td>
<td>1.0822</td>
<td>1.2133</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>1.3421</td>
<td>1.3821</td>
<td>1.0942</td>
</tr>
<tr>
<td></td>
<td><strong>MAPE 0.1051</strong></td>
<td><strong>awn</strong></td>
<td>0.2711</td>
<td>0.4920</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>1.3568</td>
<td>1.4931</td>
<td>1.5070</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>3.1731</td>
<td>4.7092</td>
<td>4.2971</td>
</tr>
<tr>
<td></td>
<td><strong>2.0350</strong></td>
<td><strong>2.0217</strong></td>
<td><strong>1.2838</strong></td>
<td><strong>2.5772</strong></td>
</tr>
<tr>
<td>T-5</td>
<td><strong>MAPE 0.0782</strong></td>
<td><strong>0.0959</strong></td>
<td><strong>0.0939</strong></td>
<td><strong>0.1196</strong></td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>3.9784</td>
<td>4.8438</td>
<td>5.9500</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>8.0391</td>
<td>13.0088</td>
<td>5.2200</td>
</tr>
</tbody>
</table>

Table (2): n = 100

<table>
<thead>
<tr>
<th>Probability distribution</th>
<th>Measures of Accuracy</th>
<th>Series mean</th>
<th>Mean of Median nearby of points(4)</th>
<th>Linear interpolation</th>
<th>Linear trend at point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>0.4540</td>
<td>0.4934</td>
<td>0.3912</td>
<td>0.4347</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>1.0057</td>
<td>1.0822</td>
<td>1.2133</td>
<td>(0.9988)</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>1.3421</td>
<td>1.3821</td>
<td>1.0942</td>
<td>(0.9231)</td>
</tr>
<tr>
<td>N (0, 1)</td>
<td><strong>MAPE 0.1051</strong></td>
<td><strong>awn</strong></td>
<td>0.2711</td>
<td>0.4920</td>
<td>0.0720</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>1.3568</td>
<td>1.4931</td>
<td>1.5070</td>
<td>1.3720</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>3.1731</td>
<td>4.7092</td>
<td>4.2971</td>
<td>3.5793</td>
</tr>
<tr>
<td>T-5</td>
<td><strong>2.0350</strong></td>
<td><strong>2.0217</strong></td>
<td><strong>1.2838</strong></td>
<td><strong>2.5772</strong></td>
<td><strong>2.0473</strong></td>
</tr>
<tr>
<td>Log-N (0, 1)</td>
<td><strong>MAPE 0.0782</strong></td>
<td><strong>0.0959</strong></td>
<td><strong>0.0939</strong></td>
<td><strong>0.1196</strong></td>
<td><strong>0.0778</strong></td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>3.9784</td>
<td>4.8438</td>
<td>5.9500</td>
<td>3.9347</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>8.0391</td>
<td>13.0088</td>
<td>5.2200</td>
<td>8.8337</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probability distribution</th>
<th>Measures of Accuracy</th>
<th>Series mean</th>
<th>Mean of Median nearby of points(4)</th>
<th>Linear interpolation</th>
<th>Linear trend at point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-N (0, 1)</td>
<td><strong>MAPE 0.0782</strong></td>
<td><strong>0.0959</strong></td>
<td><strong>0.0939</strong></td>
<td><strong>0.1196</strong></td>
<td><strong>0.0778</strong></td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>3.9784</td>
<td>4.8438</td>
<td>5.9500</td>
<td>3.9347</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>8.0391</td>
<td>13.0088</td>
<td>5.2200</td>
<td>8.8337</td>
</tr>
</tbody>
</table>
### Table (3): n = 500

<table>
<thead>
<tr>
<th>Probability distribution</th>
<th>Measures of Accuracy</th>
<th>Mean of nearby points(4)</th>
<th>Median of nearby points(4)</th>
<th>Linear interpolation</th>
<th>Linear trend at point</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (0, 1)</td>
<td>MAPE 0.8264</td>
<td>0.5845</td>
<td>0.6714</td>
<td>0.3255</td>
<td>(0.3153)</td>
</tr>
<tr>
<td></td>
<td>MAD 0.9246</td>
<td>0.9881</td>
<td>1.0223</td>
<td>1.1242</td>
<td>(0.9145)</td>
</tr>
<tr>
<td></td>
<td>MSD 1.1322</td>
<td>1.1841</td>
<td>0.9846</td>
<td>(0.6244)</td>
<td>1.3888</td>
</tr>
<tr>
<td>T-5</td>
<td>MAPE 0.1396</td>
<td>0.4179</td>
<td>0.2195</td>
<td>(0.0106)</td>
<td>0.1442</td>
</tr>
<tr>
<td></td>
<td>MAD (1.1160)</td>
<td>1.2039</td>
<td>1.1650</td>
<td>1.21714</td>
<td>1.1176</td>
</tr>
<tr>
<td></td>
<td>MSD 2.2299</td>
<td>2.4764</td>
<td>2.7124</td>
<td>(1.39526)</td>
<td>2.1924</td>
</tr>
<tr>
<td>Log-N (0, 1)</td>
<td>MAPE 1.8195</td>
<td>2.0320</td>
<td>(1.5427)</td>
<td>2.2643</td>
<td>1.8372</td>
</tr>
<tr>
<td></td>
<td>MAD (2.5730)</td>
<td>2.750</td>
<td>2.735</td>
<td>2.877</td>
<td>2.595</td>
</tr>
<tr>
<td></td>
<td>MSD (1.7398)</td>
<td>1.7693</td>
<td>2.7794</td>
<td>1.7322</td>
<td>1.7668</td>
</tr>
<tr>
<td>Binomial</td>
<td>MAPE 0.0902</td>
<td>(0.0865)</td>
<td>0.0916</td>
<td>0.1200</td>
<td>0.0904</td>
</tr>
<tr>
<td></td>
<td>MAD 4.6660</td>
<td>(4.4545)</td>
<td>4.6364</td>
<td>6.2727</td>
<td>4.6856</td>
</tr>
<tr>
<td></td>
<td>MSD 25.0020</td>
<td>24.7300</td>
<td>(20.4545)</td>
<td>105.0990</td>
<td>26.4712</td>
</tr>
</tbody>
</table>

### Table (4): n = 1000

<table>
<thead>
<tr>
<th>Probability distribution</th>
<th>Measures of Accuracy</th>
<th>Mean of nearby points(4)</th>
<th>Median of nearby points(4)</th>
<th>Linear interpolation</th>
<th>Linear trend at int</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (0, 1)</td>
<td>MAPE 0.9286</td>
<td>0.3648</td>
<td>0.3927</td>
<td>0.2340</td>
<td>(0.0536)</td>
</tr>
<tr>
<td></td>
<td>MAD (0.324)</td>
<td>0.9960</td>
<td>1.0154</td>
<td>1.057</td>
<td>0.9450</td>
</tr>
<tr>
<td></td>
<td>MSD 0.2947</td>
<td>0.2968</td>
<td>0.5864</td>
<td>(0.1182)</td>
<td>0.2873</td>
</tr>
<tr>
<td>T-5</td>
<td>MAPE 0.1113</td>
<td>0.3141</td>
<td>0.2211</td>
<td>(0.0513)</td>
<td>0.1004</td>
</tr>
<tr>
<td></td>
<td>MAD 1.1086</td>
<td>1.3334</td>
<td>1.2036</td>
<td>1.1784</td>
<td>(1.0917)</td>
</tr>
<tr>
<td></td>
<td>MSD 2.0099</td>
<td>2.7158</td>
<td>2.9257</td>
<td>(1.4329)</td>
<td>1.9919</td>
</tr>
<tr>
<td>Log-N (0, 1)</td>
<td>MAPE 2.1819</td>
<td>2.73035</td>
<td>(1.7024)</td>
<td>2.6892</td>
<td>1.9283</td>
</tr>
<tr>
<td></td>
<td>MAD (26973)</td>
<td>2.7995</td>
<td>2.9542</td>
<td>2.9998</td>
<td>2.8195</td>
</tr>
<tr>
<td></td>
<td>MSD 1.9812</td>
<td>2.0195</td>
<td>2.0019</td>
<td>(1.8998)</td>
<td>1.9796</td>
</tr>
<tr>
<td>Binomial</td>
<td>MAPE 0.0926</td>
<td>0.1209</td>
<td>0.2151</td>
<td>0.0897</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAD 4.3397</td>
<td>4.4545</td>
<td>4.6364</td>
<td>6.2727</td>
<td>4.6856</td>
</tr>
<tr>
<td></td>
<td>MSD 28.0103</td>
<td>27.0098</td>
<td>(20.4545)</td>
<td>105.0990</td>
<td>26.4712</td>
</tr>
</tbody>
</table>

In the above tables, each cell has the estimated values of PE, MAD, and MSD for each Estimation Method, and in any case (row), the best Estimation is **Bold, Italic: and in brackets.**
6 — CONCLUSION

Our simulation results demonstrate that:

I — For the symmetric distributions:

- With normal distribution, we can recommend the Linear trend at point or the linear interpolation method in most cases, especially when the sample size is large (n > 50).

- With t distribution (heavy tailed distributions):
  1 — For n < 500, we can recommend the Series mean method or the Mean of nearby points (the span of nearby points = 4).
  2 — For n > 500, we can recommend the linear interpolation method.

II - For the asymmetric distributions (log-normal distribution):

  1 — For n < 500, we can recommend the Series mean method or the Mean of nearby points (the span of nearby points = 4).
  2 — For n > 500, we can recommend the linear interpolation or the Median of nearby points method (the span of nearby points = 4).

- In general, for the continuous Distributions, we can recommend the linear interpolation method if the sample size is large. Also we can recommend the Series mean method if the sample size < 500.

III — For the Discrete Distributions (Binomial distribution):

  1 — For n < 500, we can recommend the linear trend at point or the Linear interpolation method.
  2 — For n > 500, we can recommend the Mean of nearby points or the Median of nearby points (the span of nearby points = 4).

Finally, we recommend in the future studies that changing:

1 - The sample size (n).
2 - The span of nearby points (any even or odd number).
3 — The probability distributions.
4 — The number of random samples, which we select in each case (>100).
5 - The number of missing values in the selected sample (m # 20%•n).

7 — REFERENCES

3 — Manual of SPSS - v.11 (statistical package software).
5 - Internet Websites, which are related with this paper subject.

***