DATA HANDLING AND ANALYSIS IN TECHNICAL REPORT WRITING

SAM EDO, PhD
DEPARTMENT OF ECONOMICS
UNIVERSITY OF BENIN, BENIN CITY, NIGERIA
Email: samsonedo@yahoo.com

1. **INTRODUCTION**

Data handling and analysis are very important aspects of reporting in any organization. Data have to be generated before they can be presented and analyzed. Such information/data can be obtained from two major sources.

* **Primary Source:** This involves administering questionnaires specifically designed for the purpose.

* **Secondary Source:** This involves the collection of existing data (daily, weekly, monthly, quarterly, annual, etc).

Sometimes a report employs data from both primary and secondary sources, but in most cases reports utilize data from one source. The primary source is more tedious because it requires extra efforts in designing and administering questionnaires. The questionnaires need not be superfluous, only questions that can be used to construct variables may be included. The secondary data usually cover lengthy period to avoid constraining the degree of freedom in data processing. The sources of secondary data and choice of such data are given some justification.

The data from primary and secondary sources are usually collated by sorting to remove the parts that are not relevant and arranging the relevant parts into usable form. When such data are presented in the report, the
source should be indicated for authentication, as a way of overcoming the problem of data falsification.

Data may be collected at different levels. Micro data are on economic decision-making units or institutions such as households, firms, sectors, markets, etc. Macro data results from pooling or aggregating over households, firms, sectors, markets at national level. Data collected could represent stock or flow. Stock is outcome measured at a particular point in time, while Flow is outcome measured over a period of time. Data could be qualitative or quantitative. Qualitative data can only be observed and cannot be expressed in terms of values or figures, but only involves description, e.g. colors, beauty, textures, tastes, etc. Quantitative data can be measured and expressed in terms of figures, e.g. length, volume, costs, time, size, etc. Quantitative data are more amenable to analysis.

2. TYPES OF DATA

In economic and financial analysis, three types of data may be considered. These are time series, cross section and pooled/panel data.

*Time Series Data*

These are observed values of variables overtime. The observations could be daily, weekly, monthly, quarterly, annually, quinquennially, or decennially. The data may be quantitative or qualitative. Qualitative
variables may also be called dummy variable. Time series data have stationarity problem. Time series data are stationary if their mean and variance do not vary systematically over time. If stationarity problem exist, it may be difficult to use regression estimates for purpose of forecasting.

Time series have some characteristic that could be described as trend, seasonality, cyclical, and irregular. Trend is the consistent rise or fall (on average) in the data; Seasonality is when data exhibits same behavior at corresponding periods of every year, month, week, etc; Cyclical refers to long term oscillatory movement of the mean about the trend line; Irregular describes the left over in the data upon removal of trend, seasonality and cyclical.

Time series data could also be presented as annual changes (year to year) and periodic changes of more than one year. The changes could be in levels or percentages. This is very important if the purpose of the analysis is to demonstrate growth of underlying variables. The data can also be in log if the purpose of the analysis is to get elasticity (the response of one variable to another).

*Cross-section Data*

These are observed values of variables at a given time. It could be values of GDP for several countries in a given year. Cross-section data have
the problem of heterogeneity, because some countries may have very large GDP while other countries have very small GDP. If the scale effect is not taken into account, the estimates may be biased.

**Pooled/Panel Data**

Pooled data contain both time and cross-sectional data, that is, observations for given variables covering several units and periods. Pooled data have some advantages which include large number of data points, increase in degree of freedom, efficiency of estimates, answers to some questions that cannot be addressed by time series or cross-section data, etc.

3. **DATA PRESENTATION**

The primary objective of any presentation is to convey the right message in the right form to the target audience. It should strive to get the message across accurately and effectively. In the planning of presentation, the analyst needs to identify the method that best suits the situation, which may include tables, graphs, narratives, maps, etc. If message can be communicated clearly, effectively, and with the desired impact, in a simple sentence, narrative method could be employed. If message requires precision, table containing figures and labels needs to be employed. Graphs may be used to provide visual perception of the data. Graphs may be poor means of communication, because the audience may struggle to figure out
the exact value of the data encoded in them, but some people may be impressed by such presentation.

Tables work best when the data presentation is used to look up individual values, compare individual values, requires precise values, and values involve multiple measures. Graphs work best when the data presentation is used to convey the message about the trend and shape of the data, and to show relationship between and among values.

Appropriate presentation method has some merits;
* Attracts the attention of the audience.
* Stimulates interest.
* Improves message comprehension.
* Increases retention rate and leave a lasting impression.
* Explains complex facts and processes.
* Makes abstract ideas simple and concrete.

Inappropriate presentation method has some demerits;
* Boredom.
* Alienation.
* Confusion.
* Inhibits communication.
* Dismal failure.
Tabular Presentation

This involves presentation of data in tables with rows and columns where observations are shown. A typical table has the following parts:

- Title ŭ identity of the data (what, where, when).
- Caption ŭ column title.
- Stud ŭ row title.
- Body ŭ contains the numerical information.
- Head note ŭ statement given below the title (clarifies content of the table).
- Footnote ŭ provides additional information on items in the table.
- Source ŭ where the data came from, especially secondary data.
Table 1


<table>
<thead>
<tr>
<th>Year</th>
<th>Africa</th>
<th>Asia</th>
<th>Europe*</th>
<th>Middle East</th>
<th>Western Hemisphere</th>
<th>Industrialized Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>2.5</td>
<td>5.6</td>
<td>0.2</td>
<td>8.4</td>
<td>2.3</td>
<td>2.6</td>
</tr>
<tr>
<td>1991</td>
<td>2.2</td>
<td>6.2</td>
<td>0.8</td>
<td>6.1</td>
<td>3.7</td>
<td>1.6</td>
</tr>
<tr>
<td>1992</td>
<td>0.4</td>
<td>9.0</td>
<td>1.4</td>
<td>5.0</td>
<td>2.9</td>
<td>1.8</td>
</tr>
<tr>
<td>1993</td>
<td>0.7</td>
<td>8.5</td>
<td>4.0</td>
<td>3.0</td>
<td>3.8</td>
<td>1.1</td>
</tr>
<tr>
<td>1994</td>
<td>2.6</td>
<td>9.6</td>
<td>0.5</td>
<td>2.9</td>
<td>5.1</td>
<td>3.0</td>
</tr>
<tr>
<td>1995</td>
<td>2.3</td>
<td>8.9</td>
<td>5.1</td>
<td>3.9</td>
<td>3.2</td>
<td>2.5</td>
</tr>
<tr>
<td>1996</td>
<td>6.2</td>
<td>8.0</td>
<td>4.9</td>
<td>4.5</td>
<td>3.5</td>
<td>2.9</td>
</tr>
<tr>
<td>1997</td>
<td>3.2</td>
<td>6.2</td>
<td>4.6</td>
<td>3.8</td>
<td>5.3</td>
<td>3.1</td>
</tr>
<tr>
<td>1998</td>
<td>3.6</td>
<td>3.1</td>
<td>3.3</td>
<td>3.2</td>
<td>2.2</td>
<td>2.5</td>
</tr>
<tr>
<td>1999</td>
<td>3.3</td>
<td>5.7</td>
<td>4.8</td>
<td>3.7</td>
<td>3.4</td>
<td>3.4</td>
</tr>
<tr>
<td>2000</td>
<td>2.9</td>
<td>4.5</td>
<td>5.1</td>
<td>4.1</td>
<td>3.8</td>
<td>3.2</td>
</tr>
<tr>
<td>2001</td>
<td>3.5</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>2.9</td>
<td>3.8</td>
</tr>
<tr>
<td>2002</td>
<td>3.3</td>
<td>5.3</td>
<td>3.6</td>
<td>4.3</td>
<td>3.3</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Note: * Excluding the industrialised countries in the region.
The table follows IMF regional classifications.
Diagrammatic/Graphical Presentation

This presentation could be in the form of charts and graphs.

CHART

AGRICULTURAL OUTPUT OF A WAIFEM COUNTRY

In presenting time series data, where emphasis is on comparison of values rather than how the values change over time, chart may highlight this information more effectively. But if the focus is on the changes in value overtime, graph would better serve this purpose as shown below.
4. DATA ANALYSIS

Data are sometimes analyzed by explaining and comparing values or trend in observations overtime, in which the relationship between variables may not be the objective of analysis. What is required is to depict the high periods and low periods overtime, as well as the extent of fluctuation in the trend. This analysis is quite simple and involves the plotting of graph or chart, but graph is frequently employed. An example of this type of analysis is shown in the graph above, in which deposits of commercial and merchant banks are shown to fluctuate overtime. It is clearly obvious from the graph
that changes in merchant bank deposits were larger than the changes in commercial bank deposits. The largest change in merchant bank deposit occurred about 1992, due to deregulation in the banking sector that led to emergence of more merchant banks than commercial banks.

Data are also analyzed to depict the relationships that exist between and among variables. In this case the data have to be subjected to further processing. The values obtained from this processing could also be presented in tables and graphs. For instance, money supply has impact on price level. For the quantitative assessment of this impact, we generally have to analyze past data on the variables concerned. Thus to examine the impact of money, we need to examine price level changes (inflation) associated with changes in money supply in the past periods. This can be done in two major ways - non-parametric or parametric.

Non-parametric Analysis (Analysis without prior expectations of outcomes)

This involves analysis without indicating a priori relationships. In other words, it is assumed that there is no prior knowledge of what the relationships are. This type of analysis does not have prior basis for evaluation. The variables may be represented in levels, percentage changes, log etc. Log representation is necessary if values are too large. The data may also be represented in time series or cross-section. If analysis involves only
one explanatory variable, it is quite easily done. When the analysis involves multiple explanatory variables, a line needs to be fitted, that would show approximate relationship between the dependent variable and all the explanatory variables taken together.

**Parametric Analysis (Analysis with prior expectations of outcomes)**

In the previous analysis the variables were not cast in terms of mathematical model to depict relationship based on prior notions. An appropriate model, if it is to describe the behavior of inflation must allow not only for the impact of money supply, but also for the effect of other factors, many of which may be unknown or immeasurable. The type of model usually adopted in these circumstances is called a stochastic model. If the relationship between inflation and money supply on the one hand and economic growth and money supply on the other hand is linear, then we might write our stochastic model of the two dependent variables as;

\[
\text{INF} = a_0 + a_1 \text{MS} + u, \quad a_1 > 0
\]

\[
\text{GR} = b_0 c + b_1 \text{MS} + u, \quad b_1 > 0
\]

Where; \(a_i\) and \(b_i\) are the parameters to be estimated, and \(u\) is the stochastic element (representing the effect of other factors), which is usually called the disturbance term. If the variables have linear relationship a priori, the model can be estimated using data in levels. But if they have non-linear relationship
a priori, then the model needs to be estimated in log form to minimize oscillations. If the model is estimated in log form, the estimates are taken as elasticities.

The value of the stochastic term reduces, as more variables are included in the model, as shown below;

\[ \text{INF} = a_0 + a_1 \text{MS} + a_2 \text{GE} + a_3 \text{ER} + a_4 \text{INF}_{-1} + u \]
\[ a_1 > 0, \quad a_2 > 0, \quad a_3 > 0, \quad a_4 > 0 \]

\[ \text{GR} = b_0 + b_1 \text{MS} + b_2 \text{GE} + b_3 \text{ER} + b_4 \text{GR}_{-1} + u \]
\[ b_1 > 0, \quad b_2 > 0, \quad b_3 < 0, \quad b_4 > 0 \]

Where GE is government expenditure and ER is exchange rate.

**TABLE 2**

Ordinary Least Squares Estimation Results for inflation and growth

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Explanatory variable</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>R^2</th>
<th>F-statistic</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>INF</td>
<td>INPT</td>
<td>-69.51</td>
<td>-1.05</td>
<td>0.68</td>
<td>5.76*</td>
<td>2.06</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>0.01</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GE</td>
<td>1.35</td>
<td>4.53**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER</td>
<td>0.005</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INF_{-1}</td>
<td>-4.24</td>
<td>-2.62*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>INPT</td>
<td>-28.6</td>
<td>-1.33</td>
<td>0.34</td>
<td>1.41</td>
<td>2.32</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>0.02</td>
<td>1.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GE</td>
<td>0.02</td>
<td>1.71*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER</td>
<td>-0.18</td>
<td>-1.83*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GR_{-1}</td>
<td>0.64</td>
<td>1.23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 5 percent level. **Significant at the 1 percent level
Fig. 2

Plot of Actual and Fitted Values

INFLATION

Fitted

Fig. 3

Plot of Actual and Fitted Values

GROWTH

Fitted

Years


Years

Technical Aspect of Data Analysis

When data have been presented, they are subsequently analyzed by way of interpretation and discussion. Researcher should decide the type of analysis that is ideal (concise or elaborate analysis) taking into consideration the target audience. Usually policy makers are interested in concise analysis, while academics and students would be interested in relatively more detailed analysis. Tables and figures are inserted close to explanations to enable the reader see clearly what is being explained instead of flipping through several pages to reconcile explanations with facts.

The following guidelines are important in data analysis.

* Omit aspects that are not relevant. In other words, extract and interpret only the useful parts. Analyzing all that is contained in the data could make the analysis unwieldy. Analysis therefore needs to be cleaned up to ensure tidiness.

* Emphasis should be on findings relevant to policy decisions.

* Data should be analyzed systematically, by explaining their meaning, determining whether they are consistent with expectations, and what the policy implications are.
* The tables, graphs, charts, etc., used in the discussion should as much as possible be self-explanatory. If necessary, brief notes may be appended for quick understanding.

* Avoid falsification in analysis. The tables, graphs, charts, etc., should correctly show what was obtained from data processing. Falsification could misinform and mislead readers, and inimical to policy making. Good researchers would not do that because it is unethical.

* Honesty in interpretation, by not suppressing what may appear unpleasant.

* Avoid explanation of what is not clearly understood in the processed data, or seek assistance of professional colleagues.

* Avoid inferences that are not supported by the data.

* Ensure that recommendations derive from conclusions.

**Language Aspect of Data Analysis**

In data analysis, the quality of language is very important. To ensure good quality language the following guidelines need to be followed.

- The authors need to be very clear in what they are interpreting by minimizing the use of abstract and technical terms.

- There should be precision in the use of language by selecting few words that would convey the message sharply instead of numerous
words that would cause the reader to imagine what the author intends to convey. Sentences that are too long and cumbersome should be avoided.

- Sentences should link up to ensure continuity and flow of thought; otherwise the discussions would appear like concoctions.

- Language accuracy is also important. Periods, comas, semi-colons, question marks, etc., need to be inserted in the appropriate positions.

- Language objectivity is very important. In other words, language should be devoid of bias by avoiding undue emphasis. Personal feelings or prejudice should not be allowed to control the use of language. Deficiencies, weaknesses, errors need not be covered up by use of language. The use of biased language could mislead policy makers and lead to serious policy mistakes.

**Structuring Aspect of Data Analysis**

- The analysis of data needs to the divided into sections. The sections should have linkages. In other words, each section should have a logical transition to the next one to avoid a jump/break that would cause the interpretations to be disjointed. As much as possible the number of sections should be minimized.
• Numbering: The main sections should be numbered consecutively, while the sub-sections should also carry consecutive numbers prefixed by the number of the parent section. Tables and figures should also be numbered in the same manner.

• Paragraphing: Each section or subsection should be broken into paragraphs of moderate length. Long paragraphs should be avoided, because they are not reader-friendly. The paragraphs are however justified if they seek to discuss at least one important point.

• Inserting tables and figures: They should be inserted close to where explanations are done to enable the reader see clearly what is being explained instead of flipping through several pages to reconcile explanations with facts.

• Footnotes: The use of footnotes should be minimized by adequate explanations on the body of the report. Where they cannot be avoided, such details could be provided at the bottom of relevant pages of the report. However, footnotes have become outdated as they are now only found in very few reports.