CHAPTER FIVE

Procedure for data collection and analysis

5.1 • Introduction

5.2 • Procedure for Data Collection

5.3 • Procedure for Data Analysis

5.4 • Ethical Issues

5.5 • Research Time Horizons

5.6 • Conclusion
5.1 Introduction

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes (Wikipedia- The free encyclopedia). Data collection is one of the most important stages in conducting a research. You can have the best research design in the world but if you cannot collect the required data researcher will not be able to complete his project. Data collection is a very demanding job which needs thorough planning, hard work, patience, perseverance and more to be able to complete the task successfully. Data collection starts with determining what kind of data required followed by the selection of a sample from a certain population. After that, researcher needs to use a certain instrument to collect the data from the selected sample. The techniques and procedures of data collection are the base for any type of research. Data analysis is the afterward state of data collection. Data analysis shapes up a research as it concludes the research by the help of data collected by the researcher. So it can be said that it is very important for a meaningful research findings and conclusion to be made on any research work. Hence, this section presents the sampling technique and procedure used in the collection of data and analysis in the study.

5.2 Procedure for Data Collection

This chapter contains the methods and systems adopted for the data collection. Data collection is the prelims stage of any research which should be made logically because of the entire research structure is based on the data collection. Process of data collection generally starts from the
selection of method of collecting data. Following figure is showing the entire process of data collection

**Step 1: Identify issues and/or opportunities for collecting data**

The first step is to identify issues and/or opportunities for collecting data and to decide what next steps to take. To do this, it may be helpful to conduct an internal and external assessment to understand what is happening inside and outside in sampled institution or organization.

**Step 2: Select issue(s) and/or opportunities and set goals**

The focus of Step 2 is choosing a priority issue(s) and/or opportunities for collecting data, and then setting goals and objectives.

**Step 3: Plan an approach and methods**

In Step 3, researcher will make decisions about who will be surveyed, how data will be collected, the sources of data that will be used, and the duration of the data collection project, among other questions. These decisions may be made in consultation with an expert. The methods and approaches will flow from the goals set in Step 2, and will vary significantly depending on a number of factors, including the organization’s context, size, resources, and the purpose and complexity of the issue(s) or opportunities selected.

**Step 4: Collect data**

When planning on how best to collect data in Step 4, it is important to be aware of the practical considerations and best practices for addressing logistical challenges researchers often face at this stage of the process.
**Step 5: Analyze and interpret data**

Step 5 involves analyzing and interpreting the data collected. Whether quantitative and/or qualitative methods of gathering data are used, the analysis can be complex, or less so, depending on the methods used and the amount of data collected. But according to demand of the study, research has been structured as follows:

**5.2.1 The universe and the sample**

All items in any field of inquiry constitute a “Universe” or “Population”. A complete enumeration of all items in the population is known as a census inquiry. It can be presumed that in such an inquiry, when all items are covered, no element of chance is left and highest accuracy is obtained. According to Simpson and Kafka, “A universe or population may be defined as an aggregate of items possessing a common trait or traits.”

M. Hamburg has defined it as, “The universe or population consists of the total collection of items or elements that fall within the scope of statistical investigation.”

In short, aggregate of all the items falling within the scope of enquiry is called ‘universe’. The scope of the universe depends upon the definition, object and available resources for the investigation (Prof. M.L.Oswal, 2013).

Even the slightest element of bias in such an inquiry will get larger and larger as the number of observation increases (Kothari, 2010).

Thus, the population could consist of all the persons in the country, or those in a particular geographical location, or a special ethnic or economic group, depending on the purpose and coverage of the study. A population could also consist on non-human units such as farms, houses
or business establishments. Practically population study is quite lengthy and time consuming but it generates the accurate result. In behavior population study is used where population size is relatively small. Therefore where the population size is very large population study method becomes insignificant, time and money consuming. In India, at this time government is only institution which has been carrying the population study.

Sampling is concerned with the selection of a subset of individuals, articles or commodities from within a statistical population to estimate characteristics of the whole population. This subgroup is what is referred to as the sample (Figure 5.1). It provides an alternative way of accessing the subjects and using the outcome of the data collected to make generalization on the population (O’Leary, 2005).

![Figure 5.1: Population, sample and Individual cases](source: Saunders, Lewis and Thornhill (2009: p 211))

The population in this study is consisting of two groups, namely;

1. Accounting professional in public and private sector
2. Accounting academics in Indian Universities.
The two populations were found to be appropriate in enhancing the reliability of the responses on the research instrument based on the research questions which are hereby re-stated as follows:

1. What are the common fraud prevention and detection mechanisms and their level of effectiveness in India? Is there any significant difference?
2. What are the major factors that hinder the application of forensic accounting in fraud prevention and detection in India?
3. What are the practitioners’ opinions and behavioral intention to the use of forensic accounting techniques in fraud prevention and detection in India?
4. What is the level of awareness of forensic accounting techniques among accounting practitioners and academics in India?
5. How ready are Indian universities in taking up the forensic accounting as integral part of their syllabi?

As stated the above research questions that research question 1, 2 and 3 are related to ‘the common fraud prevention and detection mechanisms, factors that hinder the use of forensic accounting services for fraud prevention and detection, and the practitioners’ opinions and behavioral intention to use forensic accounting techniques in fraud prevention and detection in India,’ would be better responded by those who are involved in the actual practice of accounting profession as chartered accountants, cost and works accountants and forensic accountants. Therefore, such a population would be constituted by accounting practitioners in the public sector as well as private sector. On the other hand, answer to question 4, which relates to the ‘the level of awareness of forensic accounting techniques among accounting practitioners, and academics in India will be provided by the accounting practitioners, and accounting academics’.
Only academics will be in the best place to respond to question 5, which deals with ‘the readiness of universities in taking up forensic accounting courses in Indian universities.’

According to Saunders, Lewis and Thornhill (2009, p. 363) “if respondents have insufficient knowledge or experience they may deliberately guess at the answer, a tendency known as uninformed response.” So an approach was adopted to reduce the possibility of uninformed response whereby two populations do not answer exactly the same question.

5.2.2 Selecting the sample

Process whereby samples are collected known as sampling. Sample is quantitative term which represent universe in the best manner. In general sense, sample is a small part of or quantity intended to show what the whole is like and sampling is the process by which this part is chosen.

With the great strides made in the theory of sampling during the past few decades, it is now possible to measure properties of mass data with calculated accuracy on the basis of samples. Consequently, today nearly all statistical surveys, whether they are for decision making in business, for policy formulation in government, or for the development of social and economic theories, are samplings. Sampling techniques that are available to any researcher can be grouped into two: probability or representative or random sampling and non-probability or judgmental sampling or purposive sampling (Saunders, Lewis and Thornhill, 2009). In probability sampling, every individual case, item or object has equal chance of being selected and the probability of it being selected is also known (Denscombe, 2007; Teddlie and Tashakkori, 2009; Saunders, Lewis and Thornhill, 2009; Udofia 2011). This thing enables researcher
to inference the result on the basis of sampling because researcher can believe on the sample that sample is properly representing the universe and answer of the questions derived on the basis of sample will be correct. However, in non-probability sampling, individual cases do not have equal opportunity of being selected and their probability of being selected is not known (Denscombe, 2007; Teddlie and Tashakkori, 2009; Saunders, Thorpe and Jackson 2009; Udofia, 2011). As such, it is impossible to answer research questions or to achieve objectives that require an estimate of the characteristics of the population by statistical method. Following figure 5.2 showing the different sampling methods

![Figure 5.2](Oswal_Agarwal_page_no. 5.8)

However, simple random sampling (SRS) is a method of selection of a sample comprising of n number of sampling units out of the population having N number of sampling units such that every sampling unit has an equal chance of being chosen. If data is heterogeneous then systematic sampling is used. For it to be a probability sampling, one may need to employ a simple random sampling techniques to choose the origin (Udofia, 2011). When the study population is heterogeneous, it is
necessary to stratify the population data before taking the samples and this technique is referred to as stratified random sampling. After stratification into sub-groups or strata, samples may be selected from each group either randomly or systematically. One advantage of stratification is that the sample is more likely going to be representative since each stratum is proportionally represented within the sample (Saunders, Lewis and Thornhill, 2009). Cluster sampling is a sampling technique used when "natural" but relatively homogeneous groupings are evident in a statistical population. In this technique, the total population is divided into these groups (or clusters) and a simple random sample of the groups is selected. Then the required information is collected from a simple random sample of the elements within each selected group. The multi stage sampling is a development of cluster sampling which can be used with geographically dispersed population to save time and money. Using all the sample elements in all the selected clusters may be prohibitively expensive or unnecessary. Under these circumstances, multistage cluster sampling becomes useful. The results of cluster sampling are based on a form of probability sampling. Stratified sampling is quite similar to cluster sampling except that groups and subgroups selected in both of the techniques are based on geographical location.

Likewise, figure 5.2 reveals that there are also four main non-probability sampling techniques. These are purposive, quota, convenience and sequential. Further, quota sampling involves selecting objects into a sample based on predetermined rule to ensure that specified numbers of objects in various sub-groups are included in the sample. In quota sampling, a population is first segmented into mutually exclusive sub-groups, just as in stratified sampling. Then judgment is used to select the subjects or units from each segment based on a specified proportion.
However, quota sampling is a completely non-random sampling and often used for survey based study. Purposive or judgmental sampling involves the selection of samples which in the opinion of the researcher are representative (Udofia, 2011) in the sense that they will provide the needed answers to the research question(s) thereby enabling the researcher to meet the objectives of the study (Saunders, Lewis and Thornhill, 2009). Judgmental sampling is a non-probability sampling technique where the researcher selects units to be sampled based on their knowledge and professional judgment. Another sampling technique which is not discussed in figure 5.2 is snowball sampling. Snowball sampling is a non-probability sampling technique where existing study subjects recruit future subjects from among their acquaintances. When it is difficult for researcher to identify other member of population then snowball sampling method is used.

Due the weaknesses of non-random sampling, the random and probability samplings were the main sampling techniques which are used in the study for the quantitative data collection. Furthermore, it is generally agreed the choice of any probability sampling technique depends on the research question and the objective of the study.

Considering the geographical spread of the population for the quantitative edge of this study (which involves the accounting academics in Indian Universities and accounting professional in public sector) and the need for a face-to-face contact with the respondents, the multi-stage sampling was used. The universities were first, stratified in two groups’ state universities and central universities as contained in the list of approved universities by the University Grant commission. (See Appendix)
However purposive sampling was adopted in selecting the central universities over the state universities. It has been validated that purposive sampling can be used when one seeks to select the cases that are particularly informative. (Neuman, 2000). The researcher belongs to university of Rajasthan which is a state university. There are 39 faculties of Accountancy and Business Statistics in university of Rajasthan. Therefore, University of Rajasthan has also selected for the purpose of sampling in addition of central universities.

<table>
<thead>
<tr>
<th>Eastern Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central University of Orissa, Orissa</td>
</tr>
<tr>
<td>Manipur University, Manipur</td>
</tr>
<tr>
<td>Mizoram University, Mizoram</td>
</tr>
<tr>
<td>Nagaland University, Nagaland</td>
</tr>
<tr>
<td>Tripura University, Tripura</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Northern Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligarh Muslim University, Uttar Pradesh</td>
</tr>
<tr>
<td>Banaras Hindu University, Uttar Pradesh</td>
</tr>
<tr>
<td>University of Allahabad, Uttar Pradesh</td>
</tr>
<tr>
<td>Rajiv Gandhi University, Arunachal Pradesh</td>
</tr>
<tr>
<td>Assam University, Assam</td>
</tr>
<tr>
<td>Tezpur University, Assam</td>
</tr>
<tr>
<td>Central University of Himachal Pradesh, Himachal Pradesh</td>
</tr>
<tr>
<td>Hemwati Nandan Bahuguna Garhwal University, Ctntal University, Uttrakhand</td>
</tr>
<tr>
<td>Central University of Haryana, Haryana</td>
</tr>
<tr>
<td>Indira Gandhi National Open University, New Delhi</td>
</tr>
<tr>
<td>Jamia Millia Islamia University, New Delhi</td>
</tr>
<tr>
<td>Jawahar Lal Nehru University, New Delhi</td>
</tr>
<tr>
<td>University of Delhi, New Delhi</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southern Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central University of Karnataka</td>
</tr>
<tr>
<td>Pondicherry University</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Western Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central University of Rajasthan</td>
</tr>
<tr>
<td>Central University of Gujarat</td>
</tr>
</tbody>
</table>

Figure 5.3 is showing established central universities all over the India regulating various accounting courses in their undergraduate and postgraduate studies. All the regions were numbered 1 to 4. This forms the first-stage sampling units from where a sample of the unit is taken using the simple random sampling technique. The sample fraction selected for this was 50 percent. Fifty percent of four is two. The strata
selected through this process are: Northern Region and Western Region. The next stage in the sampling process was to select the universities that will be finally sampled within the two region wise strata. Of the fifteen (15) central universities within the selected zones, seven (7) universities have selected. The University of Rajasthan is already fixed as a unit of sample. So there are total 8 sampling units in sampling of academics. These eight universities constitute the basic units of sampling of academics. These are: Banaras Hindu University, Uttar Pradesh, University of Allahabad, Uttar Pradesh, Jamia Milia Islamia University, New Delhi, Jawahar Lal Neharu University, New Delhi, Delhi University, New Delhi, Central University of Rajasthan, Ajmer, Central University of Himachal, Himachal and University of Rajasthan, Rajasthan.

To overcome the problem of time and the limited financial resources that is available for this research and to allow for comparability with the data obtained from the population of deductive process (academics), the states where the eight C Universities are situated were also selected for sampling the practitioners. On the basis of availability of state information, aforesaid states out of which universities are sampled, deliberate sampling was made by the researcher and following states are sampled for the sampling of practitioners.

**Sampled Units in Selected States for Practitioners**

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>Name of State</th>
<th>Number of Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Rajasthan</td>
<td>31</td>
</tr>
<tr>
<td>2.</td>
<td>Uttar Pradesh</td>
<td>81</td>
</tr>
<tr>
<td>3.</td>
<td>Delhi</td>
<td>47</td>
</tr>
<tr>
<td>4.</td>
<td>Himachal Pradesh</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>192</td>
</tr>
</tbody>
</table>
5.2.3 Nature and sources of data

The purpose of data collection is to collection of information which is beneficial for the research. Data provides the basis for beginning to address and investigate research problem which has been stated earlier in this study.

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same.

Regardless of the field of study or preference for defining data (quantitative, qualitative), accurate data collection is essential to maintaining the integrity of research. Both the selection of appropriate data collection instruments (existing, modified, or newly developed) and clearly delineated instructions for their correct use reduce the likelihood of errors occurring.

Data comes in varied forms and shapes. However, they can be classified into either primary and secondary data or quantitative and qualitative data. A common classification is based upon who collected the data. On the basis of the assumption that who collected data, data can be classified in two main categories as primary and secondary data. Primary data are such type of data which is collected by the researcher and his assistant first time. The primary data is generated by the researcher according to his need. They could be collected through interviews, questionnaires, surveys, focus groups, panels, observations and non-obstructive data
collection and set-up (Lewis, 2003). Nowadays a more privileged method of primary data collection is google forms. It is an information technology based form of questionnaire which reduces the financial cost and time consumption in research as well. For the data collection main dependence was made on the google forms in this study. Google forms can simply be described as free, collaborative survey software. It records all responses in Google Spreadsheets which is just one subset of the Google Docs applications and can be downloaded as a Microsoft Excel or Open Office file for further manipulation. To create a survey, or form, researcher needs a Google account. Researcher can then login or obtain an account at the web address of docs.google.com. In this research main tool of data collection is google forms. Structured google forms for all type of sample units as for practitioners and professionals are prepared. After the preparation of google form a short URL for each type of form has extracted and has sent to respondents’ e-mail addresses. A request call has also made to respondents to respond the enquiry form as hurry as possible. Following figures are showing the format of google forms.
Forensic accounting questionnaire for academics

Form Description

University

Gender
- Male
- Female

Age (years)
- 18-30
- 31-40
- 41-50
- 51-60
- Above 60

Education (Highest Qualification)
- U. Sc. / M. Sc. / M. Com. / M. Phil.
- PhD
- Others

Designation
- Professor
- Assistant Professor
- Associate professor
- Lecturer
- Other

Teaching Experience (in Years)
Forensic Accounting questionnaire for practitioners

Form Description

Instructions: Please tick the appropriate box or column for the option that best explains your answer.

1. Gender
   - Male
   - Female

2. Age [years]
   - 18-30
   - 31-40
   - 41-50
   - 51-60
   - 60 and above

3. Education [Highest Qualification]
   - Graduation
   - Post Graduation
   - M. Phil
   - Ph.D

4. Professional Qualification
   - CA
   - CS
   - CIA
   - MBA
   - Other
On the other hand secondary data are information that is already collected by someone else other than researcher. Data which was collected by a researcher as primary data if used by other researcher then it will be a secondary data for other one. It is implied here that every piece of work that is consulted by the researcher during the study as evidence or a support for other data source, including review of literature, is categorised as secondary data. However, Sekaran and Bougie (2010), opined that though literature review is not data gathered by the researcher for the project, it is regarded as a key component of the kind of data to be collected for the study. Saunders, Lewis and Thornhill (2007) agreed that to place literature review under secondary data does not necessarily make it to be qualitative in nature but could also be quantitative.

Secondary data may either be published data or unpublished data. Usually published data are available in: (a) various publications of the central, state are local governments; (b) various publications of foreign governments or of international bodies and their subsidiary organisations; (c) technical and trade journals; (d) books, magazines and newspapers; (e) reports and publications of various associations connected with business and industry, banks, stock exchanges, etc.; (f) reports prepared by research scholars, universities, economists, etc. in different fields; and (g) public records and statistics, historical documents, and other sources of published information. The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labor bureaus (C.R.Kothari, 2009).
5.2.4 Access Negotiation

It is known that fraud investigation is very sensitive in nature, a letter of introduction which confirmed the identity of the researcher, the institution he represents and a promise of confidentiality and anonymity was sent out to relevant authorities to enhance corporation from the study respondents. The researcher asked for help from known persons (particularly former students of Department of Accountancy and Business Statistics of the University of Rajasthan who are now practicing in many states of the India, including the ICAI, Finance Department of Ministry of Rajasthan Government, Public Sector Units) and others through telephone in arranging for the oral interviews with the respondents. For this, letter of introduction and covering letter play a vital role of entry ticket into the universities and organizations used for the study. Moreover, the same contacts persons were used for the administration of the research instrument.

During the meeting with the interview participants, formal identification of the researcher was done. The purpose for the data collection was explained in details and respondents were made to know that they were at liberty to withdraw from the exercise at any stage they did not want to continue. However, the researcher solicited their compliance to the end.

Access into these universities was negotiated through colleagues using telephones calls and e-mails. Here, the researcher formally introduced himself, his university and place of employment to these colleagues. The purpose of the research which is purely academic was also explained. In some cases, contacts were established with more than one person.
5.2.5 Sampling

Sampling was done on the two populations already discussed in section 5.2.1. These are the accounting practitioners and accounting academics in Indian universities. The actual procedures of sampling these populations are discussed as follows.

5.2.5.1 Sampling of Practitioners

Many of the accounting scholars and accounting practitioners have observed that the use of forensic accounting in fraud detection and prevention is at the organisational level. Hence, in sampling the practitioners, this research specifically focuses on the decision making process of the Chief Accounts Officers (CAO’s) in public offices (ministries and public sector units). The basis for this choice is because Chief Accounts Officers have become an integral member of the top management team involved in the organisation’s major decisions. Also, the head of the internal audit units were selected for sampling as the audit unit is supposed to be in-charge of the internal control system of the organisation. Hence, these officers are in the best position to provide the needed information. This will certainly contribute to enhancing the reliability of the data to be obtained.

The choice of government officers is based on the fact that forensic accounting is relatively a very new discipline in the world over and even newer in India. Also, India as a developing economy is heavily dependent on government. The public sector therefore is the most developed strand of the economy and most fraud that is reported in the country is perpetrated in this sector. The government coffers have become the ‘pocket’ for corruption, embezzlement and fraud. Hence, if there is any sector that should begin the use of forensic accounting techniques in
fraud prevention and detection, it is the public sector, represented in this study by government ministries and public sector units.

Nine states have been selected by the researcher with the deliberate sampling. Main object of selection of aforesaid 9 states is, these states are easily approachable for the researcher and data collection is quite easy from these states.

Table 5.2 Number of Ministries or Department in Selected States

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>Number of Ministries/Departments</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>31</td>
<td>62</td>
</tr>
<tr>
<td>2.</td>
<td>81</td>
<td>162</td>
</tr>
<tr>
<td>3.</td>
<td>47</td>
<td>94</td>
</tr>
<tr>
<td>4.</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td>Total</td>
<td>194</td>
<td>384</td>
</tr>
</tbody>
</table>

Table 5.2 shows the number of ministries and departments of government in the various States is 194. It has been stated earlier that the Chief Accounts Officer and the Head of the internal Audit in each of the ministries and departments were considered as key players in the use of forensic accounting techniques in fraud prevention and detection in the organisation, hence, these two officers were selected for sampling. This also serves to provide a check on the responses given from each establishment to ensure internal consistency in the data. The total number of samples in each unit was therefore multiplied by 2 resulting to 384 for the entire practitioners.
5.2.5.3 Sampling of academics

Population study is made for the total number of academics. Hence, the entire population served as the sample for the sampling of academics. It was therefore a census, not a sample. This is because of population of academics in the selected universities is small (89) and is distributed as shown in Table 5.3

Table 5.3: Distribution of Accounting Academics Across the Sampled Universities

<table>
<thead>
<tr>
<th>S/No.</th>
<th>University</th>
<th>No. of Academics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>University of Rajasthan, Rajasthan</td>
<td>39</td>
</tr>
<tr>
<td>2.</td>
<td>Central University of Himachal, Himachal</td>
<td>5</td>
</tr>
<tr>
<td>3.</td>
<td>Banaras Hindu University, U.P.</td>
<td>14</td>
</tr>
<tr>
<td>4.</td>
<td>University of Allahabad, U.P.</td>
<td>8</td>
</tr>
<tr>
<td>5.</td>
<td>Jamia Millia Islamia University, New Delhi</td>
<td>8</td>
</tr>
<tr>
<td>6.</td>
<td>University of Delhi, New Delhi</td>
<td>10</td>
</tr>
<tr>
<td>7.</td>
<td>Central University of Rajasthan, Rajasthan</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>89</td>
</tr>
</tbody>
</table>

5.2.7 Research instruments

This study adopts the survey strategy. An instrument is a mechanism for measuring phenomena, which used to gather and record information for assessment, decision making and ultimately understanding. An instrument such as a questionnaire is typically used to obtain factual (David Colton, 2009). The questionnaire has been described as the most suitable instrument for data collection when adopting the survey strategy in business research (Saunders, Lewis and Thornhill, 2009). Questionnaire is a document containing questions regarding research and questions are structured so as to obtain the facts. A questionnaire is a research instrument consisting of a series of questions and other prompts
for the purpose of gathering information from respondents. Questionnaire may be of many types that depend on the type of research. Following table 5.4 is showing the type and objectives of different types of questionnaire.

**Table 5.4: Types and Characteristics of questionnaires**

<table>
<thead>
<tr>
<th>Types</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured-Disguised</td>
<td>• The same question is posed to each respondent.</td>
</tr>
<tr>
<td></td>
<td>• Administering the questionnaire and post-administering work is simple i.e. coding, tabulating, etc. is easy.</td>
</tr>
<tr>
<td></td>
<td>• This type of questionnaire is least used in market research.</td>
</tr>
<tr>
<td></td>
<td>• Respondents’ biased is minimized.</td>
</tr>
<tr>
<td>Unstructured-Disguised</td>
<td>• This type of questionnaire is very commonly used for focus group discussion.</td>
</tr>
<tr>
<td></td>
<td>• This is difficult to analyze, code etc..</td>
</tr>
<tr>
<td></td>
<td>• No fixed set of questions.</td>
</tr>
<tr>
<td></td>
<td>• The inner self (why) of an individual is researched.</td>
</tr>
<tr>
<td></td>
<td>Eg: Motivation Research</td>
</tr>
<tr>
<td>Unstructured-Undisguised</td>
<td>• No fixed questions.</td>
</tr>
<tr>
<td></td>
<td>• Suitable for conducting depth interview.</td>
</tr>
<tr>
<td></td>
<td>• Subject matter can be questioned in great detail.</td>
</tr>
<tr>
<td></td>
<td>• Coding, tabulating etc. are difficult.</td>
</tr>
<tr>
<td></td>
<td>• This is not a very frequently used method.</td>
</tr>
<tr>
<td>Structured-Undisguised</td>
<td>• Fixed set of questions to every respondent.</td>
</tr>
<tr>
<td></td>
<td>• Inappropriate when researcher wants to probe deeper.</td>
</tr>
<tr>
<td></td>
<td>• Easy to administer coding, tabulating.</td>
</tr>
<tr>
<td></td>
<td>• Due to structuring and undisguised nature of the questionnaire, there is no possibility of the respondent misunderstanding the question. This is most commonly used method.</td>
</tr>
</tbody>
</table>

As discussed earlier that information used in research is sensitive nature. In such case there is none possibility of used disguised questionnaire and it is also against to ethical aspect of research. So by the keeping in mind the characteristics of questionnaires, structured and undisguised questionnaires have used by the researcher.

According to modern statistics questionnaire can be typed in two parts as follows:
According to Saunders, Lewis and Thornhill (2009), the self-administered questionnaire could be administered through the internet, postal service or delivery and collection approach and are usually completed by those making the responses. On the other hand, the interviewer-administered questionnaire could be done through telephone interview or structured interview and the responses recorded by the interviewer in line with each respondent’s answers.

Data collection through self-administered questionnaire allows for efficient means of gathering data from a large number of respondents. With this method, the respondents provide answers to the same questions. This method therefore enhances quantitative data analysis giving support to the deductive approach (Saunders, Lewis and Thornhill 2009).

There are two different survey questionnaires for the deductive method in the study. The first accounting questionnaire (appendix) was for accounting practitioners. This questionnaire has 9 sections and 76 variables. Section A named “Demographic Attributes” has 6 variables. Section B measures ‘use of fraud detection and prevention mechanisms’
with 18 variables and section C measures ‘perceived effectiveness of fraud detection and prevention mechanisms’ also with 18 variables. Section D measures “Awareness of Forensic Accounting” and comprises 5 variables while section E measures “Factors that hinder the use of forensic accounting in India” with 12 variables. Section F measured “Perceived Benefits of Forensic Accounting Techniques” with 5 items while section G has four statements which aimed to measure “Perceived Severity of Fraud in India”. Additionally, section H with 3 statements measured “Perceived Barriers of using Forensic Accounting Techniques in India” while section I has 5 items measured “Behavior towards Forensic Accounting”. The second questionnaire was for university academics. It has 2 sections with a total of 17 variables. Section, A which sought to obtain the respondents’ demographic attributes, has 6 variables. Section B measured the opinion and behavior of the respondents towards forensic accounting with 10 variables.

It must be keep in mind to researcher that response mode and questions in a questionnaire can come in different forms. The questions for the questionnaire described in this study, however, were closed-ended. The predetermined answers contained in questionnaire are hindered the respondent from to choose the answer out of scale. This kind of format is also known as the forced-response category as the respondents are forced to answer according to structured pattern. One of the most beneficial characteristics of such type of questionnaires is that such questionnaires generate frequencies of responses by the help of which researcher can enhance statistical analysis of the research. It also allows for comparisons among groups within the sample. The questions which the researcher used for the professional intention were also commonly used question. For
other questions used in this study, the researcher sought expert advice and
guidance from a team of practitioners and academics.

Except for demographic attributes and screening questions (e.g. Question: Are you aware of the existence of forensic accounting? Answer: Yes or No) which were on nominal scaling, all the other sections in the two sets of questionnaire having the ordinal and interval scales. There were also rating questions either as agreement, frequency or likelihood. Moreover, there are five and seven categories developed by Saunders, Lewis and Thornhill for four types of rating. There are some other types of rating questions developed by scholars time to time, but due to they are of amount and quantity based were not used in this research.

As Saunders, Lewis and Thornhill have argued that rating questions are frequently used with Likert-Scale. Rensis Likert developed a five-point response scale, which still bears his name (Easterby-Smith, Thorpe and Jackson, 2008) till today. The Likert scale goes from negative through neutral to positive thus it is categorized as bipolar scale. On the question that is asked to respondent, if respondent have no idea then he may allow neutral mid-point. It has two alternative response options on both the pole side of the mid-point to record moderate and extreme views for or against the issue or question. The Likert scale generally is in the following form and rank: strongly agree (5), agree (4), no opinion or undecided (3), disagree (2) and strongly disagree (1). However, there may different variant name be assigned to the different levels depending on the type of rating. For example, Saunders, Lewis and Thornhill (2009) have provided response categories for different types of rating question.

Many researchers query the rational for assigning 3 points to the level of ‘no opinion’ or ‘undecided’ and support zero (0) for it. Some other
researchers argue that five levels of Likert-scale should be reduced to four, deleting the ‘no opinion’ or ‘undecided’ completely. But, the researcher joins in the support for the assigning of zero (0) to ‘no opinion’, hence the five-point Likert-scale adopted for this study has the following ratings: strongly agree = 4, agree = 3, disagree = 2, strongly disagree = 1, indifferent or neither agree nor disagree = 0.

5.2.8 Research variables

In general sense, a variable is a measurable characteristic that varies. It may change from group to group, person to person, or even within one person over time. There are six common variable types:

1. Independent variable
2. Dependent variable
3. Intervening variables
4. Moderator variables
5. Control variables
6. Extraneous variables

Independent variables are those that the researcher has control over while dependent variables show the effect of manipulating or introducing the independent variables. Intervening variables refer to abstract processes that are not directly observable but that link the independent and dependent variables whereas moderator variables affect the relationship between the independent and dependent variables by modifying the effect of the intervening variable(s).

Control variables are those variables that are not measured in a particular study must be held constant, neutralized/balanced, or eliminated, so they will not have a biasing effect on the other variables while extraneous
variables are those factors in the research environments which may have an effect on the dependent variable(s) but which are not controlled.

In other words a variable is any observation that can take different values. The preparation of the questionnaire is made of for the fulfillment of research objectives. Most of the researcher wants to find a relationship between two sets of variables at the time of answering the questions. If it is so then researcher can divide the variables between two categories as independent variables and dependent variables. Independent variable is also known as the antecedent. On the other hand, a dependent variable is the outcome variable, the presumed effect or a variable that is caused. Dependent variable is also known as the consequent. The proper identification of variables is very important in studies that involve relationships.

Theoretical framework of a study identifies the variables involved in the study. From the framework of the study five models can be developed to show the relationship between the dependent and independent variables with the view to answering some of the research questions and testing the propositions. Table 5.5 shows the dependent and independent variables and their relationship to one another.
The kind of statistical test to be applied on the data collected is totally depend on the level or scale of measurement adopted by researcher. Thus, statistical analysis regarding selection of scale of measurement in this study, has taken very seriously. According to Easterby-Smith, Thorpe and Jackson (2008), there are two kinds of measurement levels which researchers commonly use depending on the number of distinctions between alternative points on the level of measurement.

### Table 5.5: Dependent and Independent Variables

<table>
<thead>
<tr>
<th>Models</th>
<th>Independent Variable(s)</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Educational/institutional Activities and training</td>
<td>Awareness</td>
</tr>
<tr>
<td>2.</td>
<td>Awareness</td>
<td>Perceived benefits of using forensic accounting</td>
</tr>
<tr>
<td>3.</td>
<td>Awareness</td>
<td>Perceived risks of using forensic accounting</td>
</tr>
<tr>
<td>5.</td>
<td>Perceived susceptibility to fraud Perceived severity of fraud</td>
<td>Behaviour towards forensic accounting</td>
</tr>
</tbody>
</table>
As stated in the above chart, measurement levels may be classified in two broad categories as category scales and continuous scales. Category scales are further classified into two, as unordered or nominal scale and ordered or ordinal scale. Nominal scale is simply a system of assigning number symbols to events in order to label them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales have no natural ordering while ordinal scales are naturally ordered (Easterby-Smith, Thorpe and Jackson, 2008) and involve some kind of ranking. Data that are measured on category scales are however, not amenable to parametric statistical test (Udofia, 2011).

The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule.

Continuous scales are types of ordered scales that make it possible for one to speak of more or less of whatever is being measured according to
the value on the scale (Easterby-Smith, Thorpe and Jackson, 2008). They are also classified into two, namely, interval and ratio scales. These scales allow for the use of parametric statistical tests which are the most preferred in statistical analysis because they are considered to be more powerful than the non-parametric tests (Saunders, Lewis and Thornhill, 2009). Parametric tests make use of numerical data and involve certain assumptions about the nature of the population of interest and only make sense when such assumptions hold (Blumberg, Cooper and Schindler 2008; Udofia, 2011). In the case of interval scale, which also known as continuous scale, the intervals are adjusted in terms of some rule that has been established as a basis for making the units equal.

In this study ordinal and interval scaling with Likert-scale have used for all the variables except the demographic attributes or variables, which were obtained at the nominal scale. The original Likert-scale has used in the study and all have treated as interval scale. Although it is generally observed that most textbooks on research methods and statistics treat Likert-scale questions as ordinal scale while some other author used the Likert-scale as interval scale. It is advantageous for researcher to treating ordinal data as interval data that researcher can apply parametric test because of with the non-parametric analysis can’t enhance the exploration of interesting questions. In view of the foregoing, ordinal scale data obtained in this study have treated as if they were obtained on the interval scale.
5.3 Data preparation and Analysis Procedure

5.3.1 Process of Data preparation

As above figure is showing that quantitative data should be checked, edited, coded, transcribed, cleaned and statistically adjusted before it can
be analysed (Hair 2007; Crowther and Lancaster 2009; Sekaran and Bougie 2010). However, there is no agreement on which order it should take. For this study, an initial inspection of the retrieved copies of the questionnaire was done to check for valid and invalid responses. A session of review of responded questionnaires have been done by researcher as in the google forms respondent can tick two answer for a question.

5.3.2 Data Analysis Procedure

With the data in a form that is now useful, the researcher can begin the process of analyzing the data to determine what has been learned. The method used to analyze data depends on the approach used to collect the information (secondary research, primary quantitative research or primary qualitative research). This research is a primary research the selection of method of analysis of which depends on the type of research instrument used to collect the information.

Essentially there are two types of methods of analysis – descriptive and inferential.

5.3.2.1 Descriptive Analysis

Not to be confused with descriptive research, descriptive analysis, as the name implies, is used to describe the results obtained. In most cases the results are merely used to provide a summary of what has been gathered (e.g., how many practitioners are having knowledge of Forensic Accounting) without making a statement of whether the results hold up to statistical evaluation. For quantitative data collection the most common methods used for this basic level of analysis are visual representations, such as charts and tables, and measures of central tendency including averages (i.e., mean value). For qualitative data collection, where analysis
may consist of the researcher’s own interpretation of what was learned, the information may be coded or summarized into grouping categories.

In this research bars and pie charts are used for the better presentation of data and measurements of central tendency like arithmetic mean has been used in hypothesis testing.

**5.3.2.2 Inferential Analysis**

Inferential statistics is the mathematics and logic of how this generalization from sample to population can be made. The fundamental question is: can we infer the population’s characteristics from the sample’s characteristics? Descriptive statistics remains local to the sample, describing its central tendency and variability, while inferential statistics focuses on making statements about the population.

Several inferential statistical tools were used in analysing the data and also to aid in the testing of propositions. They range between non-parametric and parametric test. Parametric tests depend on the characteristics of their population for their use; assumes that the data are normally distributed, have equal variances and thought to be robust (Ho, 2006, Udofia, 2011). On the other hand, non-parametric tests have fewer assumptions about the population, hence they often referred to as “distribution-free test” (Ho, 2006, p 357).

None of the non-parametric test used in this study. However, the parametric tests that were conducted on the data were Students’$t$-test, analysis of variance, factor analysis and correlation analysis. Their procedure and the reasons for their adoption in this study are summarised below.
The Students’ t-test or simply the t-test was used to examine whether any significance difference exist between two means. In this study t test has been used for the deciding that whether any significance difference exists between the effectiveness of fraud prevention and detection mechanisms and their actual application in fraud prevention and detection. The t-test was used because it compares the means of two sample distributions, which should be approximately normal (Uzoagulu, 1998; Ho, 2011: Udofia, 2011).

**Factor Analysis**

The major purpose of factor analysis is the orderly simplification of a large number of inter-correlated measures to a few representative constructs or factors. Suppose that a researcher wants to identify the major dimensions underlying a number of personality tests. Factor analysis is based on the assumption that all variables correlate to some degree. Consequently, those variables that share similar underlying dimensions should be highly correlated, and those variables that measure dissimilar dimensions should yield low correlations. *(Ho, 2006, p 240)*

A researcher may decide to carry out factor analysis either as a Q - mode or R- mode. A Q – mode factor analysis involves the rows resulting in the reduction of the number of observations while the R – mode deals with the columns leading to a reduction in the number of variables. It is argued that the R – mode factor analysis is the most popular since most researchers are interested in reducing the number of variables on any given research situation to a manageable size (Udofia, 2011)

There are three basic steps to factor analysis:

1. Computation of the correlation matrix for all variables
2. Extraction of initial factors

3. Rotation of the extracted factors to a terminal solution

In this study, the formation of correlation matrix containing the inter-correlation coefficients for the 12 variables was computed using the SPSS version 20 software. For this extraction of initial factor is used. There are two basic methods for obtaining factor solutions; principal components analysis and common factor analysis. However SPSS provides six methods of extraction under the common factor analysis model; these are principal-axis factoring, unweighted least-squares, generalized least-squares, maximum-likelihood, alpha factoring, and image factoring.

If the purpose is no more than to “reduce data” in order to obtain the minimum number of factors needed to represent the original set of data, then principal components analysis is appropriate. (Ho, 2006 p 242)

Since the main object of the factor analysis in this study is to know minimum number of factors that hinders the application of forensic accounting hence principal component analysis has been used in the study.

For the determination of the number of initial unrotated factors, there are two methods are existed which are traditional approached. These methods are the eigenvalues and scree test criteria. These are the eigenvalues criterion and the scree test criterion. While these two criteria are easy to use as they are Factor Analysis based on a “rules of thumb” technique that is intuitively easy to understand, they have nonetheless attracted a number of criticisms. First, the eigenvalues ≥ 1.00 rule typically overestimates, and sometimes underestimates, the number of components
extracted, and therefore does not always result in factors that are reliable (Zwick and Velicer, 1986).

Second, the scree test rule is based on a visual plot of the eigenvalues against the number of factors in their order of extraction. It involves eyeballing the plot for sharp demarcations between the eigenvalues for major and trivial factors. In practice though, such sharp demarcations do not always exist or there may be more than one demarcation point (O’Connor, 2000). Not surprisingly, the reliability of scree plot interpretation is low, even among experts (Crawford and Koopman, 1979; Streiner, 1998).

An eigenvalue is a ratio between the common (shared) variance and the specific (unique) variance explained by a specific factor extracted. Only factors with eigenvalues of 1 or greater are considered to be significant; all factors with eigenvalues less than 1 are ignored. The scree test plots the components as the X axis and the corresponding eigenvalues as the Y-axis. As one moves to the right, toward later components, the eigenvalues drop. When the drop ceases and the curve makes an elbow toward less steep decline, scree test says to drop all further components after the one starting the elbow.

But the initial factors are always difficult to interpret. Here lies the need for rotation of the extracted factors “to achieve a simpler, theoretically more meaningful factor pattern” (Ho, 2006, p 205). There are two methods: orthogonal and oblique (Ho, 2006; Udofia, 2011). While assuming that the factors are independent, orthogonal rotation process maintains the reference axes of the factors at 90º, oblique rotation allows for correlated factors in the place of independence between the rotated factors. In view of this the present study adopts the orthogonal rotation.
For adequate interpretation of factors, first of all it is important to know that data to be processed is suitable or not. One way of doing this is to visually inspect the size of factor loadings (i.e. correlation coefficients between variables and the factors they represent) on the correlation matrix. In interpreting factors, the size of the factor loadings (correlation coefficients between the variables and the factors they represent) will help in the interpretation. As a general principle, variables with large loadings indicate that they are representative of the factor, while small loadings suggest that they are not. In deciding what is large or small, a rule of thumb suggests factor loadings greater than ±0.33 are considered to meet the minimal level of practical significance. The reason for using the ±0.33 criterion is that, if the value is squared, the squared value represents the amount of the variable’s total variance accounted for by the factor. Therefore, a factor loading of 0.33 denotes that approximately 10% of the variable’s total variance is accounted for by the factor. The grouping of variables with high factor loadings should suggest what the underlying dimension is for that factor. (Ho, 2006 p 244)

**Correlation**

Correlation is primarily concerned with investigating whether a relationship exists and with determining its magnitude and direction. When two variables vary together, such as loneliness and depression, they are said to be correlated. Accordingly, correlational studies attempt to find the extent to which two or more variables are related. Typically, in a correlational study, no variables are manipulated as in an experiment—the researcher measures naturally occurring events, behaviors, or personality characteristics and then determines if the measured scores. (Ho 2006, p 219)
In order to show quantitatively the extent to which two variables are related, it is necessary to calculate a correlation coefficient. There are many types of correlation coefficients, and the decision of which one to employ with a specific set of data depends on the following factors:

- The level of measurement on which each variable is measured
- The nature of the underlying distribution (continuous or discrete)
- The characteristics of the distribution of the scores (linear or nonlinear)

Z test

A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Because of the central limit theorem, many test statistics are approximately normally distributed for large samples. For each significance level, the Z-test has a single critical value (for example, 1.96 for 5% two tailed) which makes it more convenient than the Student's t-test which has separate critical values for each sample size. Therefore, many statistical tests can be conveniently performed as approximate Z-tests if the sample size is large or the population variance known. If the population variance is unknown (and therefore has to be estimated from the sample itself) and the sample size is not large (n < 30), the Student's t-test may be more appropriate.

In this research Z test has been used for the research hypothesis 8, in which significance of difference has been checked between awareness of forensic accounting of academics and practitioners.

In both strata, number of units are more than 30 (sample size for academics is 89 and sample size for practitioners is 302), hence z test is appropriate.
5.4 Research Time Horizon

Research projects are usually done within a period (Lewis, 2003). This aspect is so important and requires attention because of the role time plays in bringing about change which research itself brings (Burns, 2005). It is also a well-known fact that time is very crucial in business (Hair et al. 2007).

Generally two time horizons are used in most of the research that is:

Cross-sectional studies

Cross-sectional studies often employ the survey strategy (Easterby-Smith et al. 2008; Robson 2002). They may be seeking to describe the incidence of a phenomenon (for example, the IT skills possessed by managers in one organisation at a given point in time) or to explain how factors are related in different organisations (e.g. the relationship between expenditure on customer care training for sales assistants and sales revenue). However, they may also use qualitative methods. Many case studies are based on interviews conducted over a short period of time.

Longitudinal studies

The main strength of longitudinal research is the capacity that it has to study change and development. Adams and Schvaneveldt (1991) point out that in observing people or events over time the researcher is able to exercise a measure of control over variables being studied, provided that they are not affected by the research process itself. One of the best-known examples of this type of research comes from outside the world of business. It is the long-running UK television series, ‘Seven Up’. This has charted the progress of a cohort of people every seven years of their life. Not only is this fascinating television, it has also provided the social
scientist with a rich source of data on which to test and develop theories of human development.

Nevertheless, the this study takes its stand on the cross-sectional study in view of the fact that time strategy is very vital to exploratory studies where the researcher contacted the respondents once for data collection, less expensive and less time consuming in comparison to the longitudinal studies (Bryrman, 2008). Also, the researcher has limited resources and time for the research.

5.5 Conclusion

This chapter of this study carefully presented the procedures that were followed by the researcher in obtaining the data for this study that are presented in the next chapter (chapter seven). It also discussed the appropriate procedures for the analyses that were performed on the data so collected in very clear terms. Samples were collected from accounting practitioners, students and academics. All procedures were adequately followed as discussed.