LESSON 23

TOOLS FOR DATA COLLECTION

Broadly there are tools of data collection as part of communication surveys. These are:

1. Interview schedule
2. Questionnaire
3. Interview Guide

As discussed earlier interview schedule and questionnaires both are predesigned list of questions used for communication with the respondents. In the case of *interview schedule*, the list of questions remains in the hands of the interviewer who asks questions from the respondent, gets his/her response, and records the responses. *Questionnaire* is also a list of questions, which is handed over to the respondent, who reads the questions and records the answers himself. For purposes of convenience *questionnaire* will refer to both interview schedule as well as questionnaire.

Interview guide is list of topics that are to be covered during the course of interview. Interview guide is used for purposes of an in-depth interviewing. Questions on the topics are formulated on the spot. Most of the questions are open ended. The interviewer may not use the same wording for each respondent; the number of questions may be different; the sequence of questions may also be different.

Guidelines for Questionnaire Design

A survey is only as good as the questions it asks. Questionnaire design is one of the most critical stages in the survey research process. While common sense and good grammar are important in question writing, more is required in the art of questionnaire design. To assume that people will understand the questions is common error. People may not simply know what is being asked. They may be unaware of topic of interest, they may confuse thee subject with something else, or the question may not mean the same thing to every respondent. Respondents may simply refuse to answer personal questions. Further, properly wording the questionnaire is crucial, as some problems may be minimized or avoided altogether if a skilled researcher composes the questions.

A good questionnaire forms an integrated whole. The researcher weaves questions together so they flow smoothly. He or she includes introductory remarks and instructions for clarification and measures each variable with one or more survey questions.

What should be asked?
The problem definition will indicate which type of information must be collected to answer the research question; different types of questions may be better at obtaining certain type of information than others.

1. Questionnaire Relevancy

A questionnaire is relevant if no unnecessary information is collected and if the information that is needed to solve the problem is obtained.
Asking the wrong or an irrelevant question is a pitfall to be avoided. If the task is to pinpoint compensation problems, for example, questions asking for general information about morale may be inappropriate. To ensure information relevancy, the researcher must be specific about data needs, and there should be a rationale for each item of information.

2. Questionnaire Accuracy

Once the researcher has decided what should be asked, the criterion of accuracy becomes of primary concern. Accuracy means that the information is reliable and valid. While experienced researchers believe that one should use simple, understandable, unbiased, unambiguous, and nonirritating words. Obtaining accurate answer from respondents is strongly influenced by the researcher's ability to design a questionnaire that facilitates recall and that will motivate the respondent to cooperate. Therefore avoid

jargon, slang, and abbreviations. The respondents may not understand some basic terminology. Respondents can probably tell thee interviewer whether they are married, single, divorced, separated, or widowed, but providing their "marital status" may present a problem. Therefore, asking somebody about his/her *marital status* while the person may not understand the meaning of marital status is likely to mess up the information. Words used in the questionnaire should be readily understandable to all respondents.

3. Avoid Ambiguity, Confusion, and Vagueness.

Ambiguity and vagueness plague most question writers. A researcher might make implicit assumptions without thinking of respondents' perspectives. For example, the question, "what is your income?" could mean weekly, monthly, or annual: family or personal; before taxes or after taxes; for this year or last year; from salary or from all sources. The confusion causes inconsistencies in how different respondents assign meaning to and answer the question.

Another source of ambiguity is the use indefinite words or response categories. Consider the words such as *often, occasionally, usually, regularly, frequently, many, good, fair,* and *poor*. Each of these words has many meanings. For one person frequent reading of *Time* magazine may be reading six or seven issues a year; for another it may be two issues a year. The word *fair* has great variety of meanings; the same is true for many indefinite words.

4. Avoid Double-Barreled Questions

Make each question about one and only one. A double barreled question consists of two or more questions joined together. It makes the respondent's answer ambiguous. For example, if asked, "Does this company have pension and health insurance benefits?" a respondent at the company with health insurance benefits only might answer either yes or no. The response has an ambiguous meaning and the researcher cannot be certain of the respondent's intentions. When multiple questions are asked in one question, thee results may be exceedingly difficult to interpret.

5. Avoid Leading Questions

Make respondents feel that all responses are legitimate. Do not let them aware of an answer that the researcher wants. A leading question is the one that leads the respondent to choose one response over another by its wording. For example, the question, "you don't smoke, do you?" leads respondents to state that they do not smoke. "Don't you think that women should be empowered?" In most the cases the respondent is likely to agree with the statement.

6. Avoid Loaded Questions

Loaded questions suggest a socially desirable answer or are emotionally charged. "Should the city government repair all the broken streets?" Most of the people are going to agree with this question simply because this is highly socially desirable. A question which may be challenging the traditionally set patterns of behavior may be considered as emotionally charged i.e. it is loaded with such material which may hit the emotions of the people. Look at some behaviors associated with masculinity in Pakistani society. Let us ask a husband "Have you ever been beaten up by your wife?" Straight away this question may be considered to be a challenge to the masculinity of the person. Hence it may be embarrassing for the person to admit such an experience. Therefore, even if the husband was beaten up by his wife, he might give a socially desirable answer.

7. Avoid Burdensome Questions that may Tax the Respondent's Memory

A simple fact of human life is that people forget. Researchers writing questions about past behavior or events should recognize that certain questions may make serious demand on the respondent's memory.

"How did you feel about your brother when you were 6 years old?" It may very difficult to recall something from the childhood.

8. Arrange Questions in a Proper Sequence

The order of question, or the question sequence, may serve several functions for the researcher. If the opening questions are interesting, simple to comprehend, and easy to answer, respondent's cooperation and involvement can be maintained throughout the questionnaire. If respondent's curiosity is not aroused at the outset, they can become disinterested and terminate the interview.

Sequencing specific questions before asking about broader issues is a common cause of question order bias. In some situations it may be advisable to ask general question before specific question to obtain the freest opinion of the respondent. This procedure, known as **funnel technique,** allows the researcher to understand the respondent's frame of reference before asking specific questions about thee level of respondent's information and intensity of his or her opinions.

9. Use Filter Question, if Needed

Asking a question that doesn't apply to the respondent or that the respondent is not qualified to answer may be irritating or may cause a biased response. Including filter question minimizes the chance of asking questions that are inapplicable. Filter question is that question which screens out respondents not qualified to answer a second question. For example the researcher wants to know about the bringing up of one's children. "How much time do you spend playing games with your oldest child?" What if the respondent is unmarried? Even if the respondent is married but does not have the child. In both these situations the question is inapplicable to him/her. Before this question the person may put a filter question whether or not the respondent is married.

10. Layout of the questionnaire

There are two format or layout issues: the overall physical layout of the questionnaire and the format of questions and responses.
Good lay out and physical attractiveness is crucial in mail, Internet, and other self-administered questionnaires. For different reason it is also important to have a good layout in questionnaires designed for personal and telephone interviews.
Give each question a number and put identifying information on questionnaire. Never cramp questions together or create a confusing appearance.
Make a cover sheet or face sheet for each, for administrative use. Put the time and date of the interview, the interviewer, the respondent identification number, and interviewer's comments and observations on it. Give interviewers and respondents instructions on the questionnaire. Print instructions in a different style from question to distinguish them.
Lay out is important for mail questionnaires because there is no friendly interviewer to interact with thee respondent. Instead the questionnaire's appearance persuades the respondents. In mail surveys, include a polite, professional cover letter on letterhead stationery, identifying the researcher and offering a telephone number for any questions. Always end with "Thank you for your participation."

LESSON 24

PILOT TESTING OF THE QUESTIONNAIRE

Pilot testing also called pre-testing means small scale trial run of a particular component; here we are referring to pilot testing of the questionnaire.

Conventional wisdom suggests that pre-testing not only is an established practice for discovering errors but also is useful for extra training the research team. Ironically, professionals who have participated in scores of studies are more likely to pretest an instrument than is a beginning researcher hurrying to complete a project. Revising questions five or more times is not unusual. Yet inexperienced researchers often underestimate the need to follow the design-test-revise process.

It is important to pilot test the instrument to ensure that the questions are understood by the respondents and there are no problems with the wording or measurement. Pilot testing involves the use of a small number of respondents to test the appropriateness of the questions and their comprehension. Usually, the draft questionnaire is tried out on a group that is selected on a convenience and that is similar in makeup to the one that ultimately will be sampled. Making a mistake with 25 or so subjects can avert the disaster of administering an invalid questionnaire to several hundred individuals. Hence the main purpose of pilot testing is to identify potential problems with the methods, logistics, and the questionnaire.

Administering a questionnaire exactly as planned in the actual study often is not possible. For example, mailing out a questionnaire might require several weeks. Pre-testing a questionnaire in this manner might provide important information on response rate, but it may not point out why questions were skipped or why respondents found certain questions ambiguous or confusing. The ability of personal interviewer to record requests for additional explanation and to register comments indicating respondent's difficulty with question sequence or other factors is the primary reason why interviewers are often used for pretest work.

What aspects to be evaluated during pilot testing?

1. Reactions of Respondents:

The reactions of the respondents can be looked at from different angles. The researcher may be familiar with the local culture; still getting the first hand experience is always useful. Going to the field, contacting the people, and their reactions to the different aspects of research may be a learning experience.

- **Availability of study population timing.** In case we are doing interviewing then pre-testing might help to find out the most appropriate time when the respondent shall be available. The researcher can plan the interviewing accordingly.
- **Acceptability of the questions asked.** An important purpose of pre-testing is to discover participants' reaction to the questions. If the participants do not find the experience stimulating when an interviewer is physically present, how will they react on the phone, or in the self administered mode? Pre-testing should help to discover where repetitiveness or redundancy is bothersome or what topics were not covered that the participant expected. An alert interviewer will look for questions or even sections that the participant perceives to be sensitive or threatening or topics about which the participant knows nothing.

- **Pre-tes**ting will also provide the opportunity to see the acceptability of the wording of the questions in the local cultural context. Some of the issues may be discussed openly while for others people use a disguised language. If people consider the use of certain phrases as offensive, then it is high time to change the wording.

- **Willingness of the respondents to co-operate.** Field testing of the questionnaire will give the idea about the level of cooperation the research team is likely to get from the respondents, particularly if they have to interview them.

2. Discovering errors in the instrument:

- **Do the tools provide you the information? Reliability. Suitability for analysis.** Tabulation of the results /of a pretest helps determine whether the questionnaire will meet the objectives of the research. A preliminary analysis often illustrates that although respondents can easily comprehend and answer a given question, it is an inappropriate question because it does not help solving the issue. The information may not be suitable for analysis.
- **Time taken/needed to interview/conduct the observation.** Pre-testing can indicate the time taken for interview or to conduct the observation. Too long questionnaires may not be recommended and, therefore, need modification. It can also help in estimating average time being taken to collect information form a respondent. Such an exercise can help in budget estimations.
- **If there is any need to revise the format of the tool.** Question arrangement can play a significant role in the success of the instrument. May be we should start with stimulating questions and place sensitive questions last. Such a situation might be handled through pre-testing. Therefore, pre-testing may help in putting questions in proper **sequence, using acceptable wording, doing appropriate translation, question spacing, structuring of answers, coding system, and needing instructions for interviewers (probing).**

3. Sampling procedure can be checked:

- **The extent to which instructions given are followed.** Field functionaries are given the instructions for following a sampling procedure. Depending upon the type of sampling to be followed, the field worker must follow the guidelines otherwise the quality of the study will be hampered. During the pre-testing one could see not only the extent to which the instructions are being followed but also locate the problems in carrying out those instructions. Also what could be the solutions to those problems?
- **How much time is needed to locate the respondents?** By following the instructions how easy it is to locate the respondents, and how much time is needed to do that activity. It could help in calculating the overall time for data collection, having relevancy for budgeting thee resources.

4. Staffing and activities of research team can be checked:

- **How successful the training has been?** Pre-testing can be seen as a period of extra training. The pre-testing exercise can provide a good opportunity to make an evaluation of the achievement of the objectives of training. For any deficiencies additional training may be provided.
- **What is the work output of each member?** The researcher can calculate the average output of each fieldworker and accordingly calculate the number of workers needed to finish the work on time. It can also help in making the budget estimates.
- **How well the research team works together?** It is a good opportunity to observe the kind of coordination the research team has. The integrated work is likely affect the efficiency of the team. Any shortcomings could be looked after.
- **Is the logistical support adequate?** Of course we are leaving the field functionaries in isolation. They shall be in need of other logistical support like the transportation, boarding, lodging, guidance and supervision. Some of these aspects could also be appraised during the pre-testing

5. Procedure for data processing and analysis can be evaluated:

- **Make dummy tables.** See how can we tabulate the data and use the appropriate statistics for purposes of interpretations

LESSON 25

INTERVIEWING

A personal interviewer administering a questionnaire door to door, a telephone interviewer calling from a central location, an observer counting pedestrians in a shopping mall, and others involved in the collection of data and the supervision of that process are all **fieldworkers.** The activities they perform vary substantially. The supervision of data collection for a mail survey differs from the data collection in an observation study. Nevertheless there are some basic issues in all kinds of fieldwork. Just for convenience, in this session we shall focus on the interviewing process conducted by personal interviewers. However, many of the issues apply to all fieldworkers, no matter what their specific setting.

Who conducts the fieldwork?

Data collection in a sponsored study is rarely carried out by the person who designs the research project. For a student, depending upon the sample size, data collection is usually done by the student himself/herself. However, the data collection stage is crucial, because the research project is no better than the data collected in the field. Therefore, it is important that the research administrator selects capable people who may be entrusted to collect the data.

There are Field Interviewing Services, who specialize in data gathering. These agencies perform door-to-door surveys, central location telephone interviewing, and other forms of fieldwork for fee. These agencies typically employ field supervisors who oversee and train interviewers, edit questionnaires completed in the field, and confirm that the interviews have been conducted.

Whether the research administrator hires in-house interviewers or selects a field interviewing service, it is desirable to have fieldworkers meet certain job requirements. Although the job requirements for different types of surveys vary, normally interviewers should be healthy, outgoing, honest, accurate, responsible, motivated, and of pleasing appearance – well groomed and properly dressed.
An essential part of the interviewing process is establishing rapport with the respondent.

In-House Training

After personnel are selected, they must be trained. The training that the interviewer will receive after being selected by a company may vary from virtually no training to one week program. Almost always there will be a **briefing session** on the particular project.

The objective of training is to ensure that the data collection instrument is administered uniformly by all fieldworkers. The goal of training session is to ensure that each respondent is provided with common information. If the data are collected in a uniform manner from all respondents, the training session will have been success.

More extensive training programs are likely to cover the following topics:

1. How to make initial contact with the respondent and secure the interview?
2. How to ask survey questions?
3. How to probe?
4. How to record responses? How to terminate the interview?

The Role of the Interviewer

Survey research interviewing is a specialized kind of interviewing. As with most interviewing, its goal is to obtain accurate information from another person.

The survey interview is a social relationship. Like other social relationships, it involves social roles, norms, and expectations. The interview is a short-term, secondary social interaction between two strangers with thee explicit purpose of one person's obtaining specific information from the other. The social roles are those of the interviewer and the interviewee or respondent. Information is obtained in a structured conversation in which the interviewer asks prearranged questions and records answers, and the respondent answers.

The role of interviewer is difficult. They obtain cooperation and build rapport, yet remain neutral and objective. They encroach on respondents' time and privacy for information that may not benefit the respondents. They try to reduce embarrassment, fear, and suspicion so that respondents feel comfortable revealing information. They explain the nature of the survey research or give hints about social roles in an interview. Good interviewers monitor the pace and direction of the social interaction as well as content of the answers and the behavior of thee respondents.

Survey interviewers are nonjudgmental and do not reveal their opinions, verbally or nonverbally. If the respondent asks for an interviewer's opinion, he or she politely redirects the respondent and indicate that such questions are inappropriate.

Stages of an Interview

Making Initial Contact and Securing the Interview

The interview proceeds through stages, beginning with introduction and entry. Interviewers are trained to make appropriate opening remarks that will convince the person that his or her cooperation is important.

Asslaam-o-Alaykum, my name is _____ and I am working for a National Survey Company. We are conducting a survey concerning "women empowerment." I would like to get a few of your ideas.

For the initial contact in a telephone interview, the introduction might be:

Asslaam-o-Alaykum, my name is _____. I am calling from Department of Social Research, Virtual University.

By indicating that telephone call is a long distance, interviewers attempt to capitalize on the fact that most people feel a long distance call is something special, unusual, or important. Giving one's personal name personalizes thee call.

Personal interviewers may carry a letter of identification that will indicate that the study is bona fide research project and not a salesman's call. The name of the research agency is used to assure the respondent that thee caller is trustworthy.

Asking the Questions

The purpose of the interview is, of course, to have the interviewer ask questions and record the respondent's answers. Training in the art of stating questions can be extremely beneficial, because interviewer bias can be a source of considerable error in survey research.

There are five major principles for asking questions:

- Ask the questions exactly as they are worded in the questionnaire.
- Read each question very slowly.
- Ask the question in the order in which they are presented in the questionnaire.

- Repeat questions that are misunderstood or misinterpreted.

Although interviewers are generally trained in these procedures, when working in thee field many interviewers do not follow them exactly. Do not take shortcuts when the task becomes monotonous. Interviewers may shorten questions or rephrase unconsciously when they rely on their memory of the question rather than reading the question as it is worded.

If the respondents do not understand a question, they will usually ask for some clarification. The recommended procedure is to repeat the question, or if the respondent does not understand a word, the interviewer should respond with "just whatever it means to you.

Often the respondents volunteer information relevant to a question that is supposed to be asked at a later point in the research. In this situation the response should be recorded under the question that deals specifically with that subject. Then rather than skip the question that was answered out of sequence, the interviewers should be trained to say something like "We have briefly discussed this, but let me ask you …." By asking every question, the interviewer can be sure that complete answers are recorded.

Probing

Probing means the verbal prompts made by field worker when the respondent must be motivated to communicate his or her answer or to enlarge on, clarify or explain an answer. Probing may be needed for two types of situations. First, it is necessary when the respondent must be motivated to enlarge on, clarify, or explain his or her answer. The interviewer must encourage the respondent to clarify or expand on answers by providing a stimulus that will not suggest the interviewer's own ideas. The ability to probe with neutral stimuli is the mark of an experienced interviewer. Second, probing may be necessary in situations in which he respondent begins to ramble or lose track of the question. In such cases thee respondent must be led to focus on specific content of the interview and to avoid irrelevant and unnecessary information. Probing is also needed when the interviewer recognizes an irrelevant or inaccurate answer.

He interviewer has several possible probing tactics to choose from, depending on the situation:

- *Repetition of the question.* The respondent who remains completely silent may not have understood the question or may not have decided how to answer it. Mere repetition may encourage the respondent to answer in such cases. For example, if the question is "What is there that you do not like about your supervisor?" and the respondent does not answer, the interviewer may probe: "just to check, is there anything you do not like about your supervisor?"
- *An expectant pause*. If the interviewer believes the respondent has more to say, the "silent probe," accompanied by an expectant look may motivate the respondent to gather his/her thoughts and give a complete response.
- *Repetition of the respondent's reply.* As the interviewer records thee response, he or she may repeat thee respondent's reply verbatim. This may stimulate the respondent to expand on the answer.
- *Neutral questions or comments.* Asking neutral question may indicate the type of information that the interviewer is seeking. For example, if the interviewer believes that thee respondent's motives should be clarified, he or she might ask, "Why do you feel that way?" If the interviewer feels that there is a need to clarify a word or phrase, then he/she might ask, "What do you mean by _____?"

Recording the Responses

The rules for recording responses to closed ended questions vary with the specific question. The general rule, however, is to place a check in the box that correctly reflects the respondent's answer.

The general instructions for recording answers to open-ended response questions is to record the answer verbatim, a task that is difficult for most people. Some of thee suggestions are:

- Record responses during the interview.
- Use the respondent's own words.
- Do not summarize or paraphrase thee respondent's answer.
- Include everything that pertains to the question objectives.
- Include all your probes.

Terminating the Interview

Fieldworkers should not close the interview before all the information has been secured. The interviewer whose departure is hasty will not be able to record those spontaneous comments respondents sometimes offer after all formal questions have been asked. Avoiding hasty departures is also a matter of courtesy.

Fieldworkers should also answer to the best of their ability any questions the respondent has concerning the nature and purpose of the study. Always leave by observing the local cultural customs. "Don't burn your bridges." Because thee fieldworker may be required to re-interview the respondent at some future time, he or she should leave thee respondent with positive feeling about having cooperated in a worthwhile undertaking. It is extremely important t thank the respondent for his or her cooperation.

The interviewer then goes to a quite and private place to edit the questionnaire and record other details such as the date, time, and place of interview; a thumbnail sketch of the respondent and interview situation, the respondent's attitude; and any unusual circumstances. The interviewer also records personal feelings and anything that was suspected.

Principles of Interviewing

The Basics

Have integrity and be honest. This is thee cornerstone of all professional inquiry, regardless of its purpose.

Have patience and tact. Interviewers ask for information from people they do not know. Thus all the rules of human relations that apply to inquiry situations – patience, tact, courtesy – apply "in spades" to interviewing.

Have attention to accuracy and detail. Among the greatest interviewing "sins" are inaccuracy and superficiality, for the professional analyst can misunderstand, and in turn mislead, a client. Do not record thee answer unless you fully understand it yourself. Probe for clarification and detailed full answers.

Exhibit a real interest in the inquiry at hand, but keep your opinions to yourself. Impartiality is imperative.

Be a good listener. Some interviewers talk too much, wasting time when respondents could be supplying more pertinent facts or opinions on the topic.

Keep the inquiry and respondents' responses confidential. Do not discuss the studies you are doing with relatives, friends, or associates. Never quote one respondent's opinion to another.

Respect others' rights. Survey research depends on the goodwill of others to provide information. There should be no coercion. Impress on prospective respondents that their cooperation is important and valuable.

Interview Bias

- Information obtained during interview should be as free as possible of bias.
- Bias could be introduced by the interviewer, interviewee, or the situation. Interviewer bias falls into six categories:

Interviewer Bias

1. Interviewer could bias the data if proper rapport is not established Errors by the respondent – forgetting, embarrassment, misunderstanding, or lying because of the presence of others.
2. Unintentional errors or interviewer sloppiness – contacting the wrong person, misreading a question, omitting questions, reading questions in the wrong order, recording wrong answer, or misunderstanding the respondent.
3. Intentional subversion by the interviewer – purposeful alteration of answers, omission or rewording of questions, or choice of an alternative respondent.
4. Influence due to the interviewer's expectations about a respondent's appearance, living situation, or other answers.
5. Failure of an interviewer to probe or to probe properly.
6. Influence on the answers due to the interviewer's appearance, tone, attitude, reactions to answers, or comments made outside of the interview schedule.

Interviewee Bias

- Errors made by the respondent –

1. Interviewees can bias the data when they do not come out with their true opinion but provide information that they think what the interviewer expects of them or would like to hear.
2. They do not understand the question, they may feel difficult or hesitant to clarify.
3. Some interviewees may be turned off because of the personal liking, or the dress of the interviewer, or the manner in which questions are put. So they may not provide truthful answers.
4. Some may provide socially undesirable answers.

Situational Bias

- Situational biases in terms of:

1. Non-participants – Unwillingness or inability to participate. Bias the sample.
2. Trust levels and rapport established by different interviewers. Elicit answers of different degrees of openness.
3. The physical setting of the interview. Respondent may not feel comfortable to be interviewed at work.

Some Tips for Interviewing

- Know the culture of the people in advance.
- Appearance – wear acceptable dress.
- Pleasantness and flexibility.
- Carry the letter of authority.
- Establish credibility and rapport. Motivating individuals to respond.
- Familiarity with the questionnaire.
- Following the question wording/ question order
- Recording responses exactly.
- Probing for responses.
- Closing the interview. No false promises. Also don't burn your bridges.
- Edit the questionnaire in the first available opportunity.

LESSON 26

SAMPLE AND SAMPLING TERMINOLOGY

A **sample** is a subset, or some part, of a larger whole. A larger whole could be anything out which sample is taken. That 'whole' could be a bucket of water, a bag of sugar, a group of organizations, a group of students, a group of customers, or a group mid-level managers in an organization. A complete group of entities sharing some common set of characteristics is population. In other words, the totality out of which sample is drawn is referred to as **population.**

Why sample?

1. Saves Cost, Labor, and Time

Applied research projects usually have budget and time constraints. Since sample can save financial cost as well as time, therefore, to go for sample study is pragmatic.
Of course, a researcher investigating a population with an extremely small number of population elements may elect to conduct a study on the total population rather than a sample because cost, labor, and time constraints are relatively insignificant.
Although sample study cuts costs, reduces labor requirements, and gathers vital information quickly, yet there could be other reasons.

2. Quality Management/supervision

Professional fieldworkers are a scarce commodity. In a large study rather than employing less qualified staff it may be advisable to do a sample study and employ highly qualified professional fieldworkers. It can certainly affect the quality of the study. At the same time it may be easier to manage a small group and produce quality information. Supervision, record keeping, training, and so forth would all be more difficult in a very large study.

3. Accurate and Reliable Results

Another major reason for sampling is that samples, if properly selected, are sufficiently accurate in most of the cases. If the elements of a population are quite similar, only a small sample is necessary to accurately portray the characteristics of interest. Most of us have had blood samples taken from the finger, the arm, or another part of body. The assumption is that blood is sufficiently similar through out the body, the characteristics of the blood can be determined on the basis of sample.
When the elements of population are highly homogenous, samples are highly representative of the population. Under these circumstances almost any sample is as good as another.
Samples may be more accurate than census. In a census study of large population there is a greater likelihood of non-sampling errors. In a survey mistakes may occur that are unrelated to the selection of people in the study. For example, a response may be coded incorrectly, or the keyboard operator might make data entry error. Interviewer mistakes, tabulation errors, and other non-sampling errors may increase during census because of the increased volume of work. In sample increased accuracy is possible because the fieldwork and tabulation of the data can be closely supervised than would be possible in a census. In field survey, a small, well trained, closely supervised group may do a more careful and accurate job of collecting information than a large group of nonprofessional interviewers trying to contact everyone.

4. Sampling may be the Only Way

Many research projects, especially those in quality control testing, require the destruction of the items being tested. If the manufacturer of firecrackers wished to find out whether each product met a specific production standard, there would be no product left after testing. Similarly, consider the case of electric

bulbs. In testing the life of bulbs, if we were to burn every bulb produced, there would be none left to sell. This is destructive sampling.

5. Determine the Period of Study

Interviewing every element of a large population without sampling requires lot of time, may be a year or more. In such a long period study, even the seasonal variation may influence the response pattern of the respondents. For example, if the study was aimed at measuring the level of unemployment in a given large city, the unemployment rate produced by the survey data would not refer to the city as of the beginning of interviewing or as of the end. Researcher may be forced to attribute the unemployment to some hypothetical date, representing to the midpoint of the study period. Hence it will be difficult to determine the exact timing to which the data of the study pertains.

Sampling Terminology

There are a number of technical terms used in books on research and statistics which need explanation. Some of the important terms are:

Element

An element is that unit about which information is collected and which provides the basis of analysis. Typically, in survey research, elements are people or certain types of people. It is that unit about which information is collected and that provides the basis of analysis. It can be a person, groups, families, organizations, corporations, communities, and so forth.

Population

A population is the theoretically specified aggregation of study elements. It is translating the abstract concept into workable concept. For example, let us look at the study of "college students." Theoretically who are the college students? They might include students registered in government colleges and/or private colleges, students of intermediate classes and/or graduate classes, students of professional colleges and/or non-professional colleges, and many other variations. In this way the pool of all available elements is population.

Target Population

Out of the conceptual variations what exactly the researcher wants to focus on. This may also be called a target population. Target population is the complete group of specific population elements relevant to the research project. Target population may also be called *survey population* i.e. that aggregation of elements from which the survey sample is actually selected.

At the outset of the sampling process, it is vitally important to carefully define the target population so the proper source from which the data are to collected can be identified. In our example of 'college students" finally we may decide to study the college students from government institutions located in Lahore, who are studying social sciences, who are aged 19 years of age, and hailing from rural areas.

Sampling

The process of using a small number of items or parts of a larger population to make conclusions about the whole population. It enables the researchers to estimate unknown characteristics of the population.

Sampling Frame

In actual practice the sample will be drawn from a list of population elements that is often different from the target population that has been defined. A sampling frame is the list of elements from which the

sample may be drawn. A simple example could be listing of all college students meeting the criteria of target population and who are enrolled on the specified date.

A sampling frame is also called the *working population* because it provides the list that can be worked with operationally. In our example, such a list could be prepared with help of the staff of the selected colleges.

Sampling Frame Error

A sampling frame error occurs when certain sample elements are excluded or when the entire population is not accurately represented in the sampling frame. The error that occurs when certain sample elements are not listed or available and are not represented in the sampling frame.

Sampling Unit

A sampling unit is that element or set of elements considered for selection in some stage of sampling. Sampling may be done in single stage or in multiple stages. In a simple, single-stage sample, the sampling units are the same as the elements. In more complex samples, however, different levels of sampling units may be employed. For example, a researcher may select a sample of *Mohallahs* in a city, and then select a sample of households from the selected *Mohallahs*, and finally may select a sample of adults from the selected households. The sampling units of these three stages of sampling are respectively *Mohallah*, households, and adults, of which thee last of these are the elements. More specifically, the terms "primary sampling units," "secondary sampling units," and "final sampling units" would be used to designate the successive stages.

Observation Unit

An observation unit, or unit of data collection, is an element or aggregation of elements from which the information is collected. Often the unit of analysis and unit of observation are the same – the individual person – but this need not be the case. Thus the researcher may interview heads of household (the observation units) to collect information about every member of the household (the unit of analysis).

Parameter

A parameter is the summary description of a given variable in a population. The mean income of all families in a city and thee age distribution of the city's population are parameters. An important part portion of survey research involves the estimation of population parameters on the basis of sample observation.

Statistic

A statistic is the summary description of a given variable in a survey sample. Thus the mean income computed from the survey sample and the age distribution of that sample are statistics. Sample statistics are used to make estimates of the population parameters.

Sampling Error

Probability sampling methods seldom, if ever, provide statistics exactly equal to the parameters that they are used to estimate. Probability theory, however, permits us to estimate the error to be expected for a given sample (more information to be sought from professional in Statistics).

LESSON 27

PROBABILITY AND NON-PROBABILITY SAMPLING

There are several alternative ways of taking a sample. The major alternative sampling plans may be grouped into probability techniques and non-probability techniques. In **probability sampling** every element in the population has a *known nonzero probability* of selection. The simple random is the best known probability sample, in which each member of the population has an equal probability of being selected. Probability sampling designs are used when the representativeness of the sample is of importance in the interest of wider generalisability. When time or other factors, rather than generalisability, become critical, non-probability sampling is generally used.

In non-probability sampling the probability of any particular element of the population being chosen is unknown. The selection of units in non-probability sampling is quite arbitrary, as researchers rely heavily on personal judgment. It should be noted that there are no appropriate statistical techniques for measuring random sampling error from a non-probability sample. Thus projecting the data beyond the sample is statistically inappropriate. Nevertheless, there are occasions when non-probability samples are best suited for the researcher's purpose.

Types of non-probability sampling:

In non-probability sampling designs, the elements in the population do not have any probabilities attached to their being chosen as sample subjects. This means that the findings from the study of the sample cannot be confidently generalized to the population. However the researchers may at times be less concerned about generalisability than obtaining some preliminary information in a quick and inexpensive way. Sometimes non-probability could be thee only way to collect the data.

Convenience Sampling

Convenience sampling (also called *haphazard or accidental sampling*) refers to sampling by obtaining units or people who are most conveniently available. For example, it may be convenient and economical to sample employees in companies in a nearby area, sample from a pool of friends and neighbors. The person-on-the street interview conducted by TV programs is another example. TV interviewers go on the street with camera and microphone to talk to few people who are convenient to interview. The people walking past a TV studio in thee middle of the day do not represent everyone (homemakers, people in the rural areas). Likewise, TV interviewers select people who look "normal" to them and avoid people who are unattractive, poor, very old, or inarticulate.

Another example of haphazard sample is that of a newspaper that asks the readers to clip a questionnaire from the paper and mail it in. Not everyone reads thee newspaper, has an interest in the topic, or will take the time to cut out the questionnaire, and mail it. Some will , and the number who do so may seem large, but the sample cannot be used to generalize accurately to the population.

Convenience samples are least reliable but normally the cheapest and easiest to conduct. Convenience sampling is most often used during the exploratory phase of a research project and is perhaps the best way of getting some basic information quickly and efficiently. Often such sample is taken to test ideas or even to gain ideas about a subject of interest.

Purposive Sampling

Depending upon the type of topic, the researcher lays down the criteria for the subjects to be included in the sample. Whoever meets that criteria could be selected in the sample. The researcher might select such cases or might provide the criteria to somebody else and leave it to his/her judgment for the actual selection of the subjects. That is why such a sample is also called as **judgmental or expert opinion sample.** For example a researcher is interested in studying students who are enrolled in a course on

research methods, are highly regular, are frequent participants in the class discussions, and often come with new ideas. The criteria has been laid down, the researcher may do this job himself/herself, or may ask the teacher of this class to select the students by using the said criteria. In the latter situation we are leaving it to the judgment of the teacher to select the subjects. Similarly we can give some criteria to the fieldworkers and leave it to their judgment to select the subjects accordingly. In a study of working women the researcher may lay down the criteria like: the lady is married, has two children, one of her child is school going age, and is living in nuclear family.

Quota Sampling

A sampling procedure that ensures that certain characteristics of a population sample will be represented to the exact extent that the researcher desires. In this case the researcher first identifies relevant categories of people (e.g. male and female; or under age 30, ages 30 to 60, over 60, etc) then decides how many to get in each category. Thus the number of people in various categories of sample is fixed. For example the researcher decides to select 5 males and 5 females under age 30, 10 males and 10 females aged 30 to 60, and 5 males and 5 females over age 60 for a 40 person sample. This is quota sampling.

Once the quota has been fixed then the researcher may use convenience sampling. The convenience sampling may introduce **bias**. For example, the field worker might select the individual according to his/her liking, who can easily be contacted, willing to be interviewed, and belong to middle class.
Quota sampling can be considered as a form of proportionate stratified sampling, in which a predetermined proportion of people are sampled from different groups, but on a convenience basis.
Speed of data collection, lower costs, and convenience are the major advantages of quota sampling compared to probability sampling. Quota sampling becomes necessary when a subset of a population is underrepresented, and may not get any representation if equal opportunity is provided to each. Although there are many problems with quota sampling, careful supervision of the data collection may provide a representative sample of the various subgroups within the population.

Snowball Sampling

Snowball sampling (also called *network, chain referral, or reputational sampling*) is a method for identifying and sampling (or selecting) cases in the network. It is based on an analogy to a snowball, which begins small but becomes larger as it is rolled on wet snow and picks up additional snow. It begins with one or a few people or cases and spreads out on the basis of links to thee initial cases.
This design has been found quite useful where respondents are difficult to identify and are best located through referral networks. In the initial stage of snowball sampling, individuals are discovered and may or may not be selected through probability methods. This group is then used to locate others who possess similar characteristics and who, in turn, identify others. The "snowball" gather subjects as it rolls along.

For example, a researcher examines friendship networks among teenagers in a community. He or she begins with three teenagers who do not know each other. Each teen names four close friends. The researcher then goes to the four friends and asks each to name four close friends, then goes to those four and does the same thing again, and so forth. Before long, a large number of people are involved. Each person in the sample is directly or indirectly tied to the original teenagers, and several people may have named the same person. The researcher eventually stops, either because no new names are given, indicating a closed network, or because the network is so large that it is at thee limit of what he or she can study.

Sequential Sampling

Sequential sampling is similar to purposive sampling with one difference. In purposive sampling, the researcher tries to find as many relevant cases as possible, until time, financial resources, or his or her

energy is exhausted. The principle is to get every possible case. In sequential sampling, a researcher continues to gather cases until the amount of new information or diversity is filled. The principle is to gather cases until a saturation point is reached. In economic terms, information is gathered, or the incremental benefit for additional cases, levels off or drops significantly. It requires that the researcher continuously evaluates all the collected cases. For example, a researcher locates and plans in-depth interviews with 60 widows over 70 years old who have been living without a spouse for 10 or more years. Depending on the researcher's purposes, getting an additional 20 widows whose life experiences, social background, and worldview differ little from the first 60 may be unnecessary.

Theoretical Sampling

In theoretical sampling, what the researcher is sampling (e.g. people, situation, events, time periods, etc.) is carefully selected, as the researcher develops grounded theory. A growing theoretical interest guides the selection of sample cases. The researcher selects cases based on new insights they may provide. For example, a field researcher may be observing a site and a group of people during week days. Theoretically, the researcher may question whether the people act the same at other times or when other aspects of site change. He or she could then sample other time periods (e.g. nights and weekends) to get more full picture and learn whether important conditions are the same.

LESSON 28

TYPES OF PROBABILITY SAMPLING

Probability samples that rely on random processes require more work than nonrandom ones. A researcher must identify specific sampling elements (e.g. persons) to include in the sample. For example, if conducting a telephone survey, the researcher needs to try to reach the specific sampled person, by calling back several times, to get an accurate sample.

Random samples are most likely to yield a sample that truly represents the population. In addition, random sampling lets a researcher statistically calculate the relationship between the sample and the population – that is the size of *sampling error.* A non-statistical definition of the sampling error is the deviation between sample result and a population parameter due to random process.

Simple Random Sample

The simple random sample is both the easiest random sample to understand and the one on which other types are modeled. In simple random sampling, a research develops an accurate sampling frame, selects elements from sampling frame according to mathematically random procedure, then locates the exact element that was selected for inclusion in the sample.

After numbering all elements in a sampling frame, the researcher uses a list of random numbers to decide which elements to select. He or she needs as many random numbers as there are elements to be sampled: for example, for a sample of 100, 100 random numbers are needed. The researcher can get random numbers from a random number table, a table of numbers chosen in a mathematically random way. Random-number tables are available in most statistics and research methods books. The numbers are generated by a pure random process so that any number has an equal probability of appearing in any position. Computer programs can also produce lists of random number.

A random starting point should be selected at the outset.

Random sampling does not guarantee that every random sample perfectly represents the population. Instead, it means that most random samples will be close to the population most of the time, and that one can calculate the probability of a particular sample being inaccurate. A researcher estimates the chance that a particular sample is off or unrepresentative by using information from the sample to estimate the sampling distribution. The sampling distribution is the key idea that lets a researcher calculate sampling error and confidence interval.

Systematic Random Sample

Systematic random sampling is simple random sampling with a short cut for random selection. Again, the first step is to number each element in the sampling frame. Instead of using a list of random numbers, researcher calculates a *sampling interval*, and the interval becomes his or her own quasi random selection method. The sampling interval (i.e. 1 in *K where K is some number)* tells the researcher how to select elements from a sampling frame by skipping elements in the frame before one for the sample.

Sampling intervals are easy to compute. We need the sample size and the population size. You can think of the sample interval as the inverse of the sampling ratio. The sampling ratio for 300 names out of 900 will be 300/900 = .333 = 33.3 percent. The sampling interval is 900/300 = 3

Begin with a random start. The easiest way to do this is to point blindly at a number from those from the beginning that are likely to be part of the sampling interval.
When the elements are organized in some kind of cycle or pattern, the systematic sampling will not give a representative sample.

Stratified Random Sample

When the population is heterogeneous, the use of simple random sample may not produce representative sample. Some of the bigger strata may get over representation while some of the small ones may entirely be eliminated. Look at the variables that are likely to affect the results, and stratify the population in such a way that each stratum becomes homogeneous group within itself. Then draw the required sample by using the table of random numbers. Hence in stratified random sampling a sub-sample is drawn utilizing simple random sampling within each stratum. (Randomization is not done for quota sampling).

There are three reasons why a researcher chooses a stratified random sample: (1) to increase a sample's statistical efficiency, (2) to provide adequate data for analyzing the various subpopulations, and (3) to enable different research methods and procedures to be used in different strata.

1. Stratification is usually more efficient statistically than simple random sampling and at worst it is equal to it. With the ideal stratification, each stratum is homogeneous internally and heterogeneous with other strata. This might occur in a sample that includes members of several distinct ethnic groups. In this instance, stratification makes a pronounced improvement in statistical efficiency.

   Stratified random sampling provides the assurance that the sample will accurately reflect the population on the basis of criterion or criteria used for stratification. This is a concern because occasionally simple random sampling yields a disproportionate number of one group or another, and the sample ends up being less representative than it could be.
   Random sampling error will be reduced with the use of stratified random sampling
   Because each group is internally homogeneous but there are comparative differences
   Between groups. More technically, a smaller standard error may result from stratified
   Sampling because the groups are adequately represented when strata are combined.

2. It is possible when the researcher wants to study the characteristics of a certain population subgroups. Thus if one wishes to draw some conclusions about activities in different classes of student body, stratified sampling would be used.
3. Stratified sampling is also called for when different methods of data collection are applied in different parts of the population. This might occur when we survey company employees at the home office with one method but mist use a different approach with employees scattered over the country.

Stratification Process

The ideal stratification would be based on the primary variable (the dependent variable) under study. The criterion is identified as an efficient basis for stratification. The criterion for stratification is that it is a characteristic of the population elements known to be related to the dependent variable or other variables of interest. The variable chosen should increase homogeneity within each stratum and increase heterogeneity between strata.
Next, for each separate subgroup or stratum, a list of population elements must be obtained. Serially number the elements within each stratum. Using a table of random numbers or some other device, a separate simple random sample is taken within each stratum. Of course the researcher must determine how large a sample must be drawn from each stratum

Proportionate versus Disproportionate

If the number of sampling units drawn from each stratum is in proportion to the relative population size of the stratum, the sample is proportionate stratified sampling. Sometime, however, a disproportionate stratified sample will be selected to ensure an adequate number of sampling units in every stratum

In a disproportionate, sample size for each stratum is not allocated in proportion to the population size, but is dictated by analytical considerations.

Cluster Sampling

The purpose of cluster sampling is to sample economically while retaining the characteristics of a probability sample. Groups or chunks of elements that, ideally, would have heterogeneity among the members within each group are chosen for study in cluster sampling. This is in contrast to choosing some elements from the population as in simple random sampling, or stratifying and then choosing members from the strata, or choosing every nth case in the population in systematic sampling. When several groups with intra-group heterogeneity and inter-group homogeneity are found, then a random sampling of the clusters or groups can ideally be done and information gathered from each of the members in the randomly chosen clusters.

Cluster samples offer more heterogeneity within groups and more homogeneity among and homogeneity within each group and heterogeneity across groups.

Cluster sampling addresses two problems: researchers lack a good sampling frame for a dispersed population and the cost to reach a sampled element is very high. A cluster is unit that contains final sampling elements but can be treated temporarily as a sampling element itself. Researcher first samples clusters, each of which contains elements, then draws a second a second sample from within the clusters selected in the first stage of sampling. In other words, the researcher randomly samples clusters, and then randomly samples elements from within the selected clusters. He or she can create a good sampling frame of clusters, even if it is impossible to create one for sampling elements. Once the researcher gets a sample of clusters, creating a sampling frame for elements within each cluster becomes more manageable. A second advantage for geographically dispersed populations is that elements within each cluster are physically closer to each other. This may produce a savings in locating or reaching each element.

A researcher draws several samples in stages in cluster sampling. In a three-stage sample, stage 1 is random sampling of big clusters; stage 2 is random sampling of small clusters within each selected big cluster; and the last stage is sampling of elements from within the sampled within the sampled small clusters. First, one randomly samples the city blocks, then households within blocks, then individuals within households. This can also be an example of **multistage area sampling.**
The unit costs of cluster sampling are much lower than those of other probability sampling designs. However, cluster sampling exposes itself to greater biases at each stage of sampling.

Double Sampling

This plan is adopted when further information is needed from a subset of the group from which some information has already been collected for the same study. A sampling design where initially a sample is used in a study to collect some preliminary information of interest, and later a sub-sample of this primary sample is used to examine the matter in more detail, is called double sampling.

What is the Appropriate Sample Design?

A researcher who must make a decision concerning the most appropriate sample design for a specific project will identify a number of sampling criteria and evaluate the relative importance of each criterion before selecting a sample design. The most common criteria

Degree of Accuracy

Selecting a representative sample is, of course, important to all researchers. However, the error may vary from project to project, especially when cost saving or another benefit may be a trade-off for reduction in accuracy.

Resources

The costs associated with the different sampling techniques vary tremendously. If the researcher's financial and human resources are restricted, this limitation of resources will eliminate certain methods. For a graduate student working on a master's thesis, conducting a national survey is almost always out of the question because of limited resources. Managers usually weigh the cost of research versus the value of information often will opt to save money by using non-probability sampling design rather than make the decision to conduct no research at all.

Advance Knowledge of the Population

Advance knowledge of population characteristics, such as the availability of lists of population members, is an important criterion. A lack of adequate list may automatically rule out any type of probability sampling..

National versus Local Project

Geographic proximity of population elements will influence sample design. When population elements are unequally distributed geographically, a cluster sampling may become more attractive.

Need for Statistical Analysis

The need for statistical projections based on the sample is often a criterion. Non-probability sampling techniques do not allow researcher to use statistical analysis to project the data beyond the sample.

LESSON 29

DATA ANALYSIS

Once the data begins to flow in, attention turns to data analysis. If the project has been done correctly, the analysis planning is already done. Back at the research design stage or at least by the completion of the proposal or the pilot test, decisions should have been made about how to analyze the data.

During the analysis stage several interrelated procedures are performed to summarize and rearrange the data. The goal of most research is to provide information. There is a difference between raw data and information.

Information refers to a body of facts that are in a format suitable for decision making, whereas **data** are simply recorded measures of certain phenomenon. The raw data collected in the field must be transformed into information that will answer the sponsor's (e.g. manager's) questions. The conversion of raw data into information requires that the data be edited and coded so that the data may be transferred to a computer or other data storage medium.

If the database is large, there are many advantages to utilizing a computer. Assuming a large database, entering the data into computer follows the coding procedure.

Editing

Occasionally, a fieldworker makes a mistake and records an improbable answer (e.g., birth year: 1843) or interviews an ineligible respondent (e.g., someone too young to qualify). Seemingly contradictory answers, such as "no" to automobile ownership but "yes" to an expenditure on automobile insurance, may appear on a questionnaire. There are many problems like these that must be dealt with before the data can be coded. Editing procedures are conducted to make the data ready for coding and transfer to data storage.

Editing is the process of checking and adjusting the data for omissions, legibility, and consistency. Editing may be differentiated from coding, which is the assignment of numerical scales or classifying symbols to previously edited data.

The purpose of editing is to ensure the completeness, consistency, and readability of the data to be transferred to data storage. The editor's task is to check for errors and omissions on the questionnaires or other data collection forms.

The editor may have to reconstruct some data. For instance, a respondent may indicate weekly income rather than monthly income, as requested on the questionnaire. The editor must convert the information to monthly data without adding any extraneous information. The editor "should bring to light all hidden values and extract all possible information from a questionnaire, while adding nothing extraneous."

Field Editing

In large projects, field supervisors are often responsible for conducting preliminary field edits. The purpose of field editing the same day as the interview is to catch technical omissions (such as a blank page), check legibility of the handwriting, and clarify responses that are logically or conceptually inconsistent. If a daily field editing is conducted, a supervisor who edits completed questionnaires will frequently be able to question the interviewers, who may be able to recall the interview well enough to correct any problems. The number of "no answers," or incomplete answers can be reduced with a rapid follow-up simulated by a field edit. The daily edit also allows fieldworkers to re-contact the respondent to fill in omissions before the situation has changed. The field edit may also indicate the need for further training of interviewers.

In-House Editing

Although almost simultaneous editing in the field is highly desirable, in many situations (particularly with mail questionnaires), early reviewing of the data is not possible. In-house editing rigorously investigates the results of data collection.

Editing for Consistency:
The in-house editor's task is to ensure that inconsistent or contradictory responses are adjusted and that answers will not be a problem for coders and keyboard punchers. Consider the situation in which a telephone interviewer has been instructed to interview only registered voters that requires voters to be 18 years old. If the editor's reviews of a questionnaire indicate that the respondent was only 17 years of age, the editor's task is to eliminate this obviously incorrect sampling unit. Thus, in this example, the editor's job is to make sure that thee sampling unit is consistent with thee objectives of the study.
Editing requires checking for logically consistent responses. The in-house editor must determine if the answers given by a respondent to one question are consistent with those given to other, related questions. Many surveys utilize filter questions or skip questions that direct the sequence of questions, depending upon respondent's answer. In some cases the respondent will have answered a sequence of questions that should not have been asked. The editor should adjust these answers, usually to "no answer' or "inapplicable," so that the responses will be consistent.

Editing for Completeness: In some cases the respondent may have answered only the second portion of a two-part question. An in-house editor may have to adjust the answers to the following question for completeness.

Does your organization have more than one Internet Web site? Yes ____ No. _____

If a respondent checked neither "yes" nor "No", but indicated three Internet Web sites, the editor may check the "yes" to ensure that this answer is not missing from the questionnaire.

Item Non-response: It is a technical term for an unanswered question on an otherwise complete questionnaire. Specific decision rules for handling this problem should be meticulously outlined in the editorial instructions. In many situations the decision rule will be to do nothing with the unanswered question: the editor merely indicates in item non response by writing a message instructing the coder to record a "missing value" or blank as the response. However, in case the response is necessary then the editor uses the **plug value.** The decision rule may to "plug in" an average or neutral value in each case of missing data. A blank response in an interval scale item with a mid point would be to assign the mid point in the scale as the response to that particular item. Another way is to assign to the item the mean value of the responses of all those who have responded to that particular item. Another choice is to give the item the mean of the responses of this particular respondent to all other questions measuring thee variables. Another decision rule may be to alternate the choice of the response categories used as plug values (e.g. "yes" the first time, "no" the second time, "yes" the third time, and so on).

The editor must also decide whether or not an entire questionnaire is "usable." When a questionnaire has too many (say 25%) answers missing, it may not be suitable for the planned data analysis. In such a situation the editor simply records thee fact that a particular incomplete questionnaire has been dropped from the sample.

Editing Questions Answered out of Order: Another situation an editor may face is thee need to rearrange the answers to an open-ended response to a question. For example, a respondent may have provided the answer to a subsequent question in his answer to an earlier open-ended response question. Because thee respondent had already clearly identified his answer, the interviewer may have avoided asking thee subsequent question. The interviewer may have wanted to avoid hearing "I have already answered that earlier" and to maintain rapport with the respondent and therefore skipped the question. To make the response appear in the same order as on other questionnaires, the editor may remove the out-of-order answer to the section related to the skipped question.

Coding

Coding involves assigning numbers or other symbols to answers so the responses can be grouped into limited number of classes or categories. The classifying of data into limited categories sacrifices some data detail but is necessary for efficient analysis. Nevertheless, it is recommended that try to keep the data in raw form so far it is possible. When the data have been entered into the computer you can always ask the computer to group and regroup the categories. In case the data have been entered in the compute in grouped form, it will not be possible to disaggregate it.

Although codes are generally considered to be numerical symbols, they are more broadly defined as the rules for interpreting, classifying, and recording data. Codes allow data to be processed in a computer. Researchers organize data into fields, records, and files. A **field** is a collection of characters (a character is a single number, letter of the alphabet, or special symbol such as the question mark) that represent a single type of data. A **record** is collection of related fields. A **file** is a collection of related records. File, records, and fields are stored on magnetic tapes, floppy disks, or hard drives.

Researchers use a coding procedure and codebook. A **coding procedure** is a set of rules stating that certain numbers are assigned to variable attributes. For example, a researchers codes males as 1 and females as 2. Each category of variable and missing information needs a code. A **codebook** is a document (i.e. one or more pages) describing the coding procedure and the location of data for variables in a format that computers can use.

When you code data, it is very important to create a well-organized, detailed codebook and make multiple copies of it. If you do not write down the details of the coding procedure, or if you misplace thee codebook, you have lost thee key to the data and may have to recode the data again.

Researchers begin thinking about a coding procedure and a codebook before they collect data. For example a survey researcher pre-codes a questionnaire before collecting thee data. Pre-coding means placing the code categories (e.g. 1 for male, 2 for female) on the questionnaire. Sometimes to reduce dependence on codebooks, researchers also place the location in the computer format on the questionnaire.

If the researcher does not pre-code, his or her first step after collecting and editing of data is to crate a codebook. He or she also gives each case an identification number to keep track of the cases. Next, the researcher transfers the information from each questionnaire into a format that computers can read.

Code Construction

When the question has a fixed-alternative (closed ended) format, the number of categories requiring codes is determined during the questionnaire design stage. The codes 8 and 9 are conventionally given to "don't know" (DK) and "no answer" (NA) respectively. However, many computer program fields recognize a blank field or a certain character symbol, such as a period (.), as indicating a missing value (no answer).

There are two basic rules for code construction. First, the coding categories should be ***exhaustive*** – that is, coding categories should be provided for all subjects or objects or responses. With a categorical variable such as sex, making categories exhaustive is not a problem. However, when the response represents a small number of subjects or when the responses might be categorized in a class not typically found, there may be a problem.

Second, the coding categories should also be ***mutually exclusive*** and ***independent***. This means that there should be no overlap between the categories, to ensure that a subject or response can be placed in only one category. This frequently requires that an "other" code category be included, so that the

categories are all inclusive and mutually exclusive. For example, managerial span of control might be coded 1, 2, 3, 4, and "5 or more." The "5 or more" category ensures everyone a place in a category.

When a questionnaire is highly structured, pre-coding of the categories typically occurs before the data are collected. In many cases, such as when researchers are using open-ended response questions, a framework for classifying responses to questions cannot be established before data collection. This situation requires some careful thought concerning the determination of categories after editing process has been completed. This is called post-coding or simply *coding.* The purpose of **coding open-ended response questions** is to reduce the large number of individual responses to a few general categories of answers that can be assigned numerical scores. Code construction in these situations necessarily must reflect the judgment of the researcher. A major objective in code-building process is to accurately transfer the meaning from written answers to numeric codes.

Code Book

A book identifying each variable in a study and its position in thee data matrix. The book is used to identify a variable's description, code name, and field. Here is a sample:

Q/V No.	Field/ col. No.	Code values
--	1-5	Study number
-	6	City
		1 = Lahore
		2 = Rawalpindi
		3 = Karachi
	7 -9	Interview No.
Sex	10	1 = Male
		2 = Female
Age	11-12	Actual
Education	13	1 = Non literate
		2 = Literate

Production Coding

Transferring the data from the questionnaire or data collection form after the data have been collected is called production coding. Depending upon the nature of the data collection form, codes may be written directly on the instrument or on a special coding sheet.

Data Entries

Use of scanner sheets for data collection may facilitate the entry of the responses directly into the computer without manual keying in the data. In studies involving highly structured paper questionnaires, an **Optical scanning system** may be used to read material directly to the computer's memory into the computer's memory. Optical scanners process the marked-sensed questionnaires and store thee answers in a file.

Cleaning Data

The final stage in the coding process is the error checking and verification, or **"data cleaning"** stage, which is a check to make sure that all codes are legitimate. Accuracy is extremely important when coding data. Errors made when coding or entering data into a computer threaten the validity of measures and cause misleading results. A researcher who has perfect sample, perfect measures, and no errors in gathering data, but who makes errors in the coding process or in entering data into a computer, can ruin a whole research project.