PAKISTAN NETHERLANDS PROJECT ON HUMAN RESOURCES

Third Training Course on Labour Information Processing

lecture notes on survey methods

Lecture 1: Introduction

The following topics will be dealt with in this lecture:
- What is a sample survey
- The need for sample surveys in the field of human resources
- The limitations of sample surveys
- The need for clear objectives
- The information production process
- Steps involved in survey design
- Planning of the survey

A sample survey is a way to gather information. I will not try to give a precise definition, but broadly one may say that on the base of surveying a relatively small number of subjects, you try to find general conclusions for the entire population.

The purpose of surveying is of course to provide data to somebody who can later use it for any purpose. We often use it for planning purposes, but one can also think of commercial purposes etc.

The type of data required depends on the type of planning (sectorial, geographic etc.)

The need for sample surveys arises when not all information can be found in statistical sources or administrative records. In Human resources, for example, there is an evident need for more information on the informal sector, which can not or to a limited extend be found in Annual Establishment Enquiries and Labour Force Surveys.

The limitations of sample surveys are mainly in the enormous amount of resources they are absorbing. In other words, a sample survey is a very expensive way of collecting information. Therefore, it should only be undertaken if no other source to collect the information is available.

Another limitation of surveys is that they are done by human beings. It has so many stages where things can go wrong that it almost requires a superman to avoid the errors. A good surveyor must be a statistician, an artist, a linguist, a good manager, an economist (costs) and a psychologist. I hope to show you later during our discussions at what stage you need what type of skill.

In your jobs you have to produce information. Maybe in the form of ready processed data, in the form of a report, or in the form of financial accounting etc. Anyhow, all this is production of information.

In fact, without thinking too much, the production of this information requires several steps. We often do not know this, as we are involved in one or two steps only. This is often the case in big organizations.

Let us have a closer look at these different stages of the production of information.

1) Definition of required information
2) Planning and preparation
3) Data collection
4) Data processing
5) Reporting

In this seminar, we will focus on collecting and production of information by means of sample surveys. In figure 1.1, the main steps in survey design are presented. It is an elaboration on the rather simple scheme for information production as above.

[about here figure 1.1]

1) Determination of objectives, this means:
- A detailed dialogue with the user
- Special requirements must be clear
- You must know what is the use of the data
- The motivation of the user must be clear
- At what level of precision the data must be gathered
- User requirements vs. producer constraints

2) Major issues affecting the survey include:
- Geographic and socio-economic conditions of the region
- Transport and communication facilities
- Availability of manpower and equipment

3) Major decisions to be taken
- Once or repetitive
- One or more main subjects for research
- Scope/coverage of survey
- Survey period
- Method of data collection (mail, personal interviews, administrative records, direct observation etc.
- Concepts and definitions
- Unit of investigation

The scope and coverage of the survey include a precise definition of the population. Usually, the survey purpose will define the survey population broadly in terms of geographical extent. However, there are always marginal groups, homeless etc., and it depends on the purpose of the survey whether they may be excluded.

concepts and definitions are very important:
- They must be perfectly clear.
- They should fit local conditions, but preferably take into account and be comparable with international classifications.
- If you do not make your concepts and definitions clear, you will not be able to formulate precise questions to find the information.

It is very important to choose the right unit of investigation
For example: compare the following findings by two different surveys:
- In 40% of firms workers have no contract
- 98% of workers do have a contract
The results are consistent if 40% of firms are small scale and employ only 2% of workers without contract.

PRELIMINARY PLAN

Planning is essential for a good survey. Information which is delayed can be useless. The most beautiful reports are worthless if they are not written in time. Planning may save a lot of time.

Example of a field of planning:

The field organization plan should include:
- Recruitment
- Preparation of manuals
- Training of survey officers
- Quality control

For all the steps in survey design plans must be made in terms of resources and time.

PRETEST AND PILOT SURVEY
Five basic objectives of pilot surveys are:
1. To provide the means of testing the survey forms and the sampling techniques. Are there enough answer categories on the forms?
2. For training the enumerators.
3. To provide some information on the variability of the population within the study domain. Is the stratification used really decreasing the sampling error?
4. To provide a basis for estimating the costs and times required for the entire survey. The duration of interviews and the time taken for interviewers to travel from one place to another can rarely be determined accurately beforehand.
5. To determine the most effective type and size of sampling unit.

The sampling technique for pilot surveys should be identical to the sampling technique used for the entire survey. However, to estimate variances in the strata with a very small sample size may be contradictory. There is no satisfactory solution for this, except that one can sample only from the most important strata in the pilot survey. Specific problems of other strata will not be found then.

REVISIONS:
- Sample design
- Questionnaire
- Field organization
- Tabulation programme

DATA COLLECTION
- Most expensive part of survey, so do it good at once.
* Requires cars, people and organization.
* Do not start if previous steps are not yet ready.
* Special lecture is devoted to data collection

DATA PROCESSING
- Preparation, editing and coding (special lecture)
- Processing
- Analyzing

REPORTING
- Good report is essential for the user. It must save him time.

What are the main objectives of this course?
1) To make you enthusiastic on surveys
2) To provide you with some insight what steps are required to perform a survey
3) To give you some practical experience with surveying

What are we not going to do
1) Providing an extensive theoretical knowledge on sample designs and statistical methods. There are many handbooks, also available in Islamabad, which provide good summaries of the different methods.

Assignment:
1) Define the objective and target group for a small sample survey in Islamabad.
2) Make a time planning of one of the resources required to perform the survey.
3) Give the outline of the main tables you would like to obtain from this survey.

Lecture 2: Sampling methods 1

In this lecture, we will discuss:
- Concepts and definitions
- Statistical concepts
- Nominal, ordinal, interval and ratio variables
- Applications of statistical methods in surveying

See reader on statistical methods for sample surveys

Assignment: p. 8 statistical methods manual

Lecture 3: Sampling methods 2

In this lecture, we will discuss what is needed to draw a sample from a larger population. Three things are important; you must know how big your sample must be, where you can take it from and how you will draw the sample. In survey terminology the following words are used:

- Sample frames
- Sample size
- Sample designs

Suppose that we want to estimate the average income of workers in Pakistan with the help of a survey. How do sample frame, size and design influence the quality of our estimates?

There are two concepts which help to measure the quality:
-1 Bias from the mean.
-2 Precision of our estimation.

We do not know the real figure of income of the entire population. What we may know is that if we repeat the sampling several times, then we may on average approach the real figure.
This depends however on the:
- Sample design (sample may not be representative, non-random sampling is used etc.).
- Non-sampling errors (wrong questions etc.).

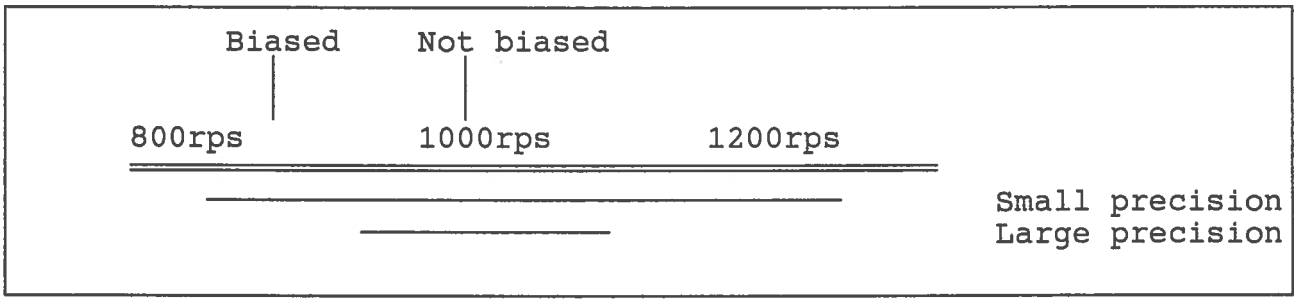

Figure 3.1: Bias and precision in case real average wage Rps. 1000

The second thing influencing the quality is the size of the range around our estimated mean which includes with, say, 95% confidence, the real mean. In statistics we call this the 95% confidence interval. If this is a large range, we have a small precision. If this is a small range, the precision is large. The range is often expressed in standard deviations.

Both sample size and sample design influence the precision of the estimated income.

Knowing this, please indicate the best type of sample for estimating average income in Pakistan:
1) 2 million people in the town only or 700 over the entire country
2) 2 million people chosen by enumerator or 700 in a random way

SAMPLE FRAMES

Sample frames are lists, records or maps from which you will have to select the sample. For example, for the AEE there is an outdated list of firms with 20 employees and more which are registered.

The problems which may appear when using a sample frame are:
1) Missing elements
The frame can be inadequate in the sense that it does not cover the entire population you want to survey. The frame can be incomplete when some part of the population which is supposed to be on the frame is not. In these cases, you must change the coverage of the survey or use additional frames.

2) Clusters of elements
If there are, for example, more individuals living at one address and your unit of investigation is the individual, you may chose to: -take all the elements in one cluster or -select one individual from the cluster. We will discuss this during cluster sampling.

3) Frame is not up to date.
For example, people on the list may have died or emigrated. Draw a larger sample when this is the case in the pilot survey.

4) Duplicated elements
If the same unit is more than once represented on the sample frame, it may be necessary to reweigh the elements on the frame.

SAMPLE SIZE

The sample size is only important if there are no significant sampling and non sampling errors and if the sample design is good. For a given design, an increase of the sample size will increase the precision, but will not eliminate the bias.

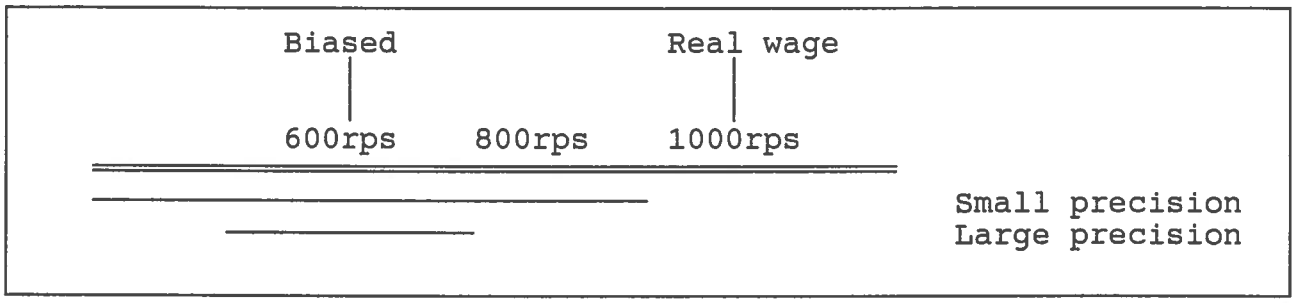

Figure 3.2: Bias and precision in case real average wage Rps. 1000 and biased estimate of Rps. 600

Suppose we have no bias from the mean. How big must be our sample?

Basically, there are three approaches
1) Cost approach, depending on how much money is available
2) Tentative approach, often based on experience
3) Statistical approach

In the last approach, the policy maker must determine two things:
1) Size of the maximum allowed error (MAE)
2) Degree of confidence that the bias will not exceed the MAE

The size of the maximum allowed error is the difference between our estimated mean x and the real mean μ.

$$MAE = x - \mu$$

The degree of confidence is expressed in the number, z, of standard deviations of the sample mean, x is away from μ.
z = 1,96 relates to a 95% confidence interval
In formula form:

$$x - \mu = z * \sigma x$$

We know from statistical theory that the standard deviation of the sample means, σx, is approximately equal to the standard deviation of the population, σ, divided by the square root of the sample size,

$$\sigma x = \sigma / \sqrt{n}$$

If we combine the three formulas, we obtain,

$$MAE = z * \sigma / \sqrt{n}$$

Rearranging gives,

$$n = z^2 * \sigma^2 / MAE^2$$

Problems:
- σ is often not known. It may be found from past surveys.
- More variables in the survey are important (more σ's relevant).
- A breakdown of results is necessary, e.g. income per sector of activity.

So, the help of the user is necessary for the statistician!

It is very important to adjust for the non-response rate when calculating the required sample size.

SAMPLE DESIGNS

There are many ways of drawing a sample from the population. Some of these ways replace the unit of investigation back on the sample frame after being selected: Sampling with replacement. We will restrict ourselves, however, to sampling without replacement, or simple random sampling, a more precise method with less statistical complications.

Some of the sample designs are sound, others are not. One of the characteristics of sound sample designs is that all of them use random sampling in one way or another. Remember that if sampling is done by non-random methods, the selection is almost always (un)consciously influenced by human choice.

Some sample designs of special importance are:
- Simple random sampling
- Systematic sampling
- Stratification
- Cluster sampling
- Multi-stage sampling
- Sampling with PPS
- Area sampling
- Multiphase sampling

Replicated sampling and quota sampling are important sample designs which fall outside the scope of this seminar.

- Simple random sampling is a method in which the units of investigation are selected randomly from the sampling frame. Often tables with random numbers are used for this. The method is like a lottery in which numbers are taken from a bowl.

- Systematic sampling
As long as there is absolutely no order in the sampling frame, systematic sampling can be used. In this method, with N units on the frame and n units to be selected, a random number k between 0 and N/n is selected. This is the first unit on the list to be selected. After this number k, every N/n number is being selec-

ted. However, any order in the list may have terrible consequences. For example, on a list with only couples, there is a good chance that only males are selected, which could result in a serious bias.

- Stratification

To increase the precision of the survey, you do not only have to increase the size of the survey. Stratification is another method. Before any selection is made, the population is divided in a number of strata. Within the strata, the selection is done randomly.

Example, suppose you are to select two firms out of four to estimate the average number of workers.

Firm	Workers	Strata
1	2	A
2	4	A
3	10	B
4	12	B

The average number of workers is 28/4=7. If we select randomly two firms, there is a good chance that we will select firms 1 and 2 with on average 3 workers or firms 3 and 4 with on average 11 workers. Both results are far from reality.

If we select one firm from strata A (small firms) and one firm from strata B (big firms), the possible range of sample means is from 6 to 8 only. The precision of our sample is increasing with stratification (decreasing σx).

The advantages of stratification are that:

- Proportionate stratification will never lead to worse results as compared to simple random sampling.
- If the user is interested in results by separate strata, he is at least sure that there are sufficient units in this strata.
- It may be administratively easy to divide the population and the work.
- Many sampling frames are already subdivided by groups or areas.
- Different parts of the population may call for different sampling procedures. For example, big firms are often on lists of registration. Small firms are not. Big firms should maybe be approached in a different way as compared to small firms.

Of course, there cannot be more strata than units. In general, the number of units must be at least twice the number of strata.

- Cluster sampling

Just as in the case of stratification, the population is divided in clusters or strata. However, with cluster sampling there is a random selection of clusters from which a 100% sample is taken. No unit should be in more than one cluster! This method is often cheaper and easier to apply in underdeveloped areas where there are rarely satisfactory sampling frames.

The cluster is not always an area. It can also be a household from which all members are selected. Every entity which contains a number of units to be investigated can be called a cluster.

The precision of clustering is almost always lower as compared to simple random sampling. This is the result of intra class correlation. People living in a particular district or town almost always have some characteristics which make them look more like each other than like people in other districts.

Two remarks:
1) The more heterogeneous clusters are, the less precision is lost by clustering.
2) It is better to have many small clusters than a few large ones.

- Multi-stage sampling

In practice it is often necessary to sample in more stages. First, primary sampling units (PSU) are selected. A sample is taken from the PSUs. Secondary sampling units are listed. A sample is taken from SSUs.

The advantage is that the SSUs only have to be prepared and listed from the PSUs in the first sample selection.

PSUs in Pakistan are often districts.

Suppose, we have a suitable list of districts and of individuals living within the districts. We would like to draw a sample of 2,000 adult civilians. We can look at four possibilities:

1) Single random sampling:
Select 2,000 adults from the list of adults for the entire country.

2) Unstratified two-stage sample
Sample 1 would be too thinly spread over the country. Concentrate sample in, say, 100 districts, taking 20 people from each district.

3) Stratified simple random sampling
Stratify country by region and take adults from each region according to the population share. Note that here you also need a list of adults for the entire country.

4) Stratified, two-stage sampling
Same as in 3), but with selection of SSUs (administrative districts) from each region.

- Sampling with PPS

If PSUs such as districts differ too much in size, then it is wrong to take 1 out of 10 districts by means of simple random sampling followed by taking a fixed number of units from the district. The population in larger districts would have a smaller chance of being selected as compared to the population units in smaller districts. The selection would not be random anymore.

The solution to this problem is that if one PSU has twice as large a population as another, it should be given twice the chance of being selected. If the same number of persons is then selected from each of the PSUs chosen, the overall probability of selection of any person will be the same.

PPS sampling has two advantages:
1) It has a larger precision than random sampling of PSUs. With random selection a large PSU can easily be excluded and when it is included it can have a heavy weight.
2) It is easier for fieldwork, as all enumerators have the same number of interviews to be done.

One limitation for sampling with PPS is that the size of PSU must be known or estimated in advance.

- Area sampling

Area sampling is the main method of sampling without adequate lists of population or firms. The area to be covered by the survey (country, province or town) is divided in a number of smaller areas. Area sampling is basically multi stage sampling in which maps rather than lists or registers serve as the sampling frame.

In area sampling, it is important to have the boundaries well defined. Streets, railways and rivers often make good boundaries.

Stratification and PPS sampling are often used together with area sampling.

There are different ways of selecting dwellings from an area:

In the city, often with the help of a map, blocks are sampled. The blocks provide at the same time information for geographic stratification. When the blocks have been defined, numbered, allocated to strata and assigned measures of size, a stratified PPS sample of them can be drawn. Dwellings can then be listed from which a sample will be taken, or sub-blocks in the block can be defined from which all dwellings are taken. Each area in the survey should be defined in such a way that it coincides with those areas for which statistical data are available.

In rural areas usually more stages are necessary. In rural area there is often a subdivision to administrative areas, townships, chunks, and blocks. Especially at the lower end often fieldwork must be done to register all blocks etc.

- Multiphase sampling

In multiphase sampling some units are asked for more information than others. This method is often used if there are less important detailed matters for which less precision is required or in the case that some information is very costly to acquire.

If the large sample is done first, it can often be used to stratify the smaller sample. Also the non-response rate can then be better estimated for the smaller sample.

Assignments:
1) Make a sample design for the Islamabad survey.
2) Try to find out what types of sample frames are available.

Reading material:
Moser and Kalton, Survey Methods for Social Investigation, ch. 5 to 7

Lecture 4: Questionnaire design

This lecture deals with:
- The art of questionnaire design
- Wording of questions
- Format of questionnaire
- Attitudes, behaviour and consciousness of respondent

No survey can be better than its questionnaire. The discussion about the questionnaire should start at the early planning stages and finish only after the completion and revision of pilot surveys.

What are the features of a good questionnaire?
- Lay out and printing are such that it is easy for the person who must fill it in.
- Lay out and printing are such that it is easy for the data processor.
- It covers the problem definition of the survey.
- It has not too many questions.
- It obeys to the rules of wording and question content.
- It takes the respondent into account.

All these aspects will be discussed below. First more technical, how is a questionnaire built up? Second, more respondent-oriented, what do we know about the respondent?

Everybody understands that words go into questions and that questions go into questionnaires. What many people do not understand is that writing questions does not give you a questionnaire. A questionnaire is a structure consisting of different layers or parts. Each of the parts must be considered individually, but at the same time also integrated into the final shape to form a complete questionnaire.

Different elements in a questionnaire are:

- Length
- Format
- Words
- Questions

LENGTH

Restrict the number of questions. A questionnaire easily grows from a short list of questions to a full document. It is better to restrict oneself to the problem to be analysed.

Remember that a long questionnaire:
- is costly in terms of resources.

- may be competitive with the number of interviews or, in other words, will restrict the sample size and therefore also the reliability of the survey.
- is demoralizing for respondents and enumerators and may therefore lead to higher refusal rates.

FORMAT of questionnaire:

1) Sequence of questions is very important: do not ask respondent to evaluate safety of nuclear plants if you first ask questions about radioactive waste, nuclear accidents etc. Rather start from general questions and only then proceed with more detailed and personal questions. Non-response will be lower.

2) Convenient for interviewer.
Many questionnaire designers never had to do the field work themselves. Referring to other points in questionnaire should be avoided as much as possible. With a trained enumerator doing the job, the form should be focussed on efficiency of field handling rather than on attractiveness. It can be more codified. Examples of more codified questionnaires are those used in the Labour Force Matrix, the Annual Establishment Enquiry etc.. These fact finding surveys do not really include questions, but often use an extensive code-book for the enumerator.

3) If the respondent has to fill in the questionnaire, the form must be kept simple. Use must be made of bold and capital letters to emphasize some words or instructions. Questions should be simple and not too many.

4) Convenient for editors and coders so that they can process and introduce the data smoothly. However, the enumerator has priority.

WORDS

Question wording per se has little impact on the stability of results. However, as soon as the variations in wording introduce a different concept or emotion surrounding the issue, the difference can be substantial.

Avoid:
- technical terms
- complicated language
- words which may have different meanings for different groups
- long questions with more meaningful words
- ambiguous words (you in English can be plural as well as singular)
- Vague words (regular, often, kind of)

Questionnaires should not be a political football. If you use the word family planning rather than birth control when asking people whether they restrict the number of children they take, the real restriction of children taken will be underestimated. Many people think family planning is something as budgeting for the future or even regard it as some type of financial management.

QUESTIONS

Two main issues:
1) The type of question and
2) The quality of the question (good or bad)

Ad 1) Types of questions
- facts questions: (what is the day today)
- opinion questions (how do you like Fridays)
- knowledge questions (where does the name Friday originate from)
- motivation questions (why do you never work on Friday)
for all of these: open ended or closed questions are possible.

With facts questions the enumerator may change the wording in order to find out whether the respondent knows the fact. With opinion questions this is never allowed. Opinion questions are difficult to interpret, because:
- answers are many-sided.
- no correct answer exists.
- they depend on the mood or what was seen on television yesterday.
- people have different involvement in the topic.
- they are difficult to check against other data of the questionnaire.

For example, the question "are you in favour or against death punishment" can be considered from different angles.
Morally: you shall not kill
Religious: it is allowed from religious law
Legal: not according to the law
Economic: less expensive than keeping person in prison
Emotional: my sister was killed, so the murderer should be punished by death

Note that no correct answer exists. If there was recently a program on television about murderers escaping from prison, more people may be in favour of death punishment.

Ad 2) The quality of questions

Of course, even good questions depend on the willingness and ability of respondents to answer the question. People may exaggerate even if they are willing/able to answer. For example,

a person who tells he is director in the bank may be the man who is directing people to the appropriate window.

Bad questions are those that are:
1) Not understandable for respondent, because wording or concepts are too difficult.
2) Unanswerable, because respondent can not have the knowledge or in answer categories one type of answer is missing.
3) Leading in that respondent is forced into an answer that he should otherwise not give.
4) Not specific enough.
5) Not direct enough (can you tell me....).

Example of questions that are <u>not specific</u> enough

A survey investigates the satisfaction of customers with the prices of meals and the service in a specific restaurant. Consider the following questions:
1) Are you satisfied with this restaurant?
2) Are you satisfied with the service in this restaurant?
3) Are you satisfied with the prices of meals?
4) Are you satisfied with the price of butter in this restaurant?

Ad1) Not specific enough, no distinction between price and service
Ad2) Not specific enough, there are many aspects of service
Ad3) Not specific enough, there may be special meal prices
Ad4) Very specific, but may lead to too many questions

The solution in this case is to do research before the survey in order to find out about different meals, prices, service etc.

When people answer questions, the answer may not be satisfactory for the researcher for the following reasons:
1. Bias: question wording that makes one answer more likely than the other.
2. Non validity: the question does not really measure the concept to be investigated.
3. Non reliability: you cannot repeat the experiment with the same results.
4. Not meaningful: Answers can be reliable, unbiased and valid, but still meaningless. why?:

Ad 4)
a) Respondent can lie.
b) Due to the type of question the respondent cannot tell what he really feels etc.
c) Non consciousness of respondent.
d) We may misunderstand the truth telling respondent, because he gives us only part of the total picture.
e) Lack of knowledge of respondent makes that more knowledge would change his attitude and, hence, answer.

f) In the case we impute meaningfullness on respondent

In some cases, pre coded answer categories increase the response. When asking about income, more respondents will be willing to answer if they have to indicate their income group, rather than their precise income.

MORE ABOUT THE RESPONDENT

What is necessary to know or expect about the respondent?

- Relation between attitude and behaviour respondent.
- Level of consciousness of respondent.
- Influence of environment on attitude/behaviour respondent.
- Knowledge respondent may reasonably be assumed to have.
- Meaning of words for respondent, multidimensionality of problem.
- Importance of past behaviour of respondent.

ATTITUDE AND BEHAVIOUR

You would expect that a persons attitudes and beliefs determine his behaviour. However, if you ask people what is the most important problem the nation is facing nowadays, it is most often something which received most attention in the press! In this case behaviour (reading the newspaper) determines attitude.

Moreover, people often say and think different than what they do. Their behaviour differs from their attitude. People who say that they take care about the environment still go to work by means of a private car.

CONSCIOUSNESS

Let me start to say that both in Pakistan and Netherlands many people are mostly unconscious. People often do not know or understand why they do things. The more unconscious, the larger the gap between what people say and what they do.

Estimating the level of consciousness should be done by focussing on the following questions:
- Does the respondent know what he <u>feels</u> or is he simply answering the questions?
- On what <u>experience</u> are answers based?
- Does the respondent see the <u>implications</u> of what he says?
- Does he see the whole problem or only <u>parts</u> of it?
- Is he <u>inconsistent</u> in answering questions?

Direct experience leads to more consciousness. Therefore you can include experience related questions in survey. The example given before in the case of small scale enterpreneurs is to follow the

question whether he would like to have any credit by the question whether he has ever applied for credit before.

The responses from many surveys support the conclusion that people know what they do, but do not know why they do it. In other words, they are not really conscious. In the case of "suppose that" questions, people usually cannot grasp the different pieces of information, put them together and then imagine what their behaviour should be.

People are driving many kilometers for obtaining cheaper products without taking the costs of petrol into account.

Lack of consciousness is often indicated by inconsistent answers, not based on objective reality, and purely emotional. The more unconsciousness in answers, the more difficult it will be to predict behaviour:

Example: If you ask whether a small scale entrepreneur wants to have a bank credit and most of them have no experience at all with credits, then even when they say yes, then it does not mean that they will really apply if it is available.

ENVIRONMENT

Until man is conscious, he is manipulated by the environment. The environment is personal (health, age, sex, race, status etc.) and spatial (car, home, town, state etc.)

If respondents are unconscious, they are more likely to be influenced by their environment. They are reactive instead of active. Many answers can only be understood in the context of respondent environment.

Reading material:

Labaw, Patricia, J., Advanced Questionnaire Design, Abt Books, Cambridge, Massachusetts, 1980

Moser and Kalton, Survey Methods for Social Investigation, ch. 13

Assignment:
1) Design a one-page questionnaire for the Islamabad survey. Do not forget the coding.

This form is the property of:

© SOCIAL SURVEYS (GALLUP POLL) LIMITED 1969
211 Regent Street, London, W.1.

THE GALLUP POLL
CQ 658 November 1969

Private and Confidential

11/12 Interviewer's No.

Contact No.

Ask the questions exactly as they are worded. DO NOT READ OUT THE ANSWERS unless indicated.

13 Thinking back over the last 7 days have you discussed with members of your family or friends, or thought about, any of these things? (Show CARD X and mark all mentioned.)

1	2	3	4	5	6	7	8	9
10	20	30	40	50	60	70	80	90
100	200	300	400		–V None of these			

16a) If there were a General Election tomorrow, which party would you support?

1 Conservative
2 Labour
3 Liberal
4 Nationalist*
5 Other
–6 Don't know

b) *If don't know:* Which would you be most inclined to vote for?

7 Conservative
8 Labour
9 Liberal
0 Nationalist*
X Other
–V Don't know

*Welsh or Scottish

ASK ALL

17 Were you able to go and vote in the General Election in March, 1966 or were you prevented?

–1 No, did not vote

Yes, voted for:
2 Conservative
3 Labour
4 Liberal
5 Other

Ask all who voted: For which candidate did you vote?

20 How important would you say these difficulties are in the way of Britain joining the Common Market:

	Very Important	Important	Not very Important	Not at all Important	Don't know
a) The difficulties facing our agriculture if the Government subsidies have to be withdrawn?	1	2	3	4	R
b) The difficulties for the Commonwealth if Britain has to end the preference which she now gives to Commonwealth goods?	5	6	7	8	R
c) The difficulty of getting the Common Market to meet the problems of the other European countries now with Britain in the European Free Trade Area?	9	0	X	V	R

21 What do you think would be the advantages to Britain of joining the Common Market? Any other advantages or arguments *for* joining the Common Market? (Write in fully below and ring all appropriate codes)

1 Keep down the cost of living, cheaper goods
2 British industries/agriculture will benefit/expand
3 Full employment will be maintained
4 Better wages/working conditions
5 Social/medical/educational services will improve
6 Good (in the long run) for the Commonwealth
7 Fewer travel restrictions in ECM countries
8 Wider range of goods in the shops
9 Closer *political* links with European countries
0 Closer *military* links with European countries
X Other advantages
V No advantages at all
R Don't know

ASK ALL

18 Now that British membership of the Common Market or European Community is again in the news, here are some issues which will have to be settled between Britain and the members of the Common Market.

Taking each one separately, do you think it would be a good thing or a bad thing for Britain to agree to: (Read out and mark for each. For every other contact begin at (h) and work upwards.)

	Good	Bad	Don't know
a) Gradual removal of all trade barriers between members	1	2	3
b) The removal eventually of all subsidies and other protection to industries, like agriculture	4	5	6
c) Unemployment policies and social services to be brought more into line with each other	7	8	9
d) Inviting workers here from other member countries, to take up jobs which cannot be filled by our own labour	0	X	V
e) Adopting the principle of "equal pay for equal work" as between men and women	1	2	3
f) Pooling our resources to develop atomic energy for peaceful uses	4	5	6
g) Supporting a common tariff for customs barrier against all countries not in the Common Market	7	8	9
h) To have as the goal achieving closer political relationships	0	X	V

22 And what would be the disadvantages of joining the Common Market? Any other disadvantages or arguments *against* joining the Common Market? (Write in fully below and ring all appropriate codes)

1 Cost of living, prices will go up
2 British industries/agriculture will suffer
3 Employment in Britain will be affected, labour imported from Europe
4 Cut in wages/working conditions
5 Increased competition for British industry
6 Social/medical/educational services will suffer
7 Commonwealth will be affected
8 Loss of political identity/sovereignty
9 Too close *political* links with European countries
0 Too close *military* links with European countries
X Other disadvantages
V No disadvantages at all
R Don't know

23. If you never work on the computer, please indicate the reason(s):

	Do not fill in
o No computers available	1
o Not enough training facilities	2
o Other tasks on the job take toomuch time	3
o Nobody can handle the computer	4
o Other (please specify):.....	5

Importance of subjects course for the present job

24. Please indicate the importance of the following course subjects for your PRESENT WORK. We would like to know to what degree you are using the knowledge which you gained during the course

	5 Very Important	4 Important	3 Less important	2 Not important	1 Don't know
Planning studies	o	o	o	o	o
Sample surveys	o	o	o	o	o
Lotus 123	o	o	o	o	o
WordPerfect	o	o	o	o	o
Statistical sources	o	o	o	o	o
Guest lectures	o	o	o	o	o

Lecture 5: Data collection

In this lecture, we will deal with:
- Methods of data collection.
- Attitude and capabilities of field enumerators.
- Tasks of field enumerators.
- Supervisors.
- Administration of field work.
- Problems in field work.
- How to deal with non-response.

METHODS OF DATA COLLECTION

Different methods of data collection are:
* Field interview
* Mail questionnaire
* Telephone interview
* Administrative records
* Direct observation

We will concentrate on field interviews, i.e. the case where trained staff visits places where units are located and where they make the required observation.

Units can be persons, households, products, trees, markets. Observation can be made by interview, counting etc.

ATTITUDE AND CAPABILITIES OF FIELD ENUMERATORS

The capacities of the field worker must include:
* Clearness of understanding
* Desire for accuracy
* No bias/prejudices
* Neat handwriting
* Tactful
* Patient
* Honest
* Responsible
* Healthy
* Active
* Confidential

Hence, the enumerator answering these requirements is the perfect human being. Interviewing usually requires working in the evening as well as during the day and can be both mentally and physically exhausting. The quota for each interviewer to complete in one day should therefore not be too great and tailored towards his capabilities.

TASKS OF FIELD ENUMERATORS

1)
The enumerator must receive full instructions how to locate and identify units, how to read maps, and what to do with unnamed streets, unnumbered dwellings, new buildings, buildings disappeared etc.

The enumerator should also know the meaning of concepts and definitions used by the survey as only he can sometimes locate them in the field. For example, in a survey of retail shops, the enumerator should know whether this includes street vendors, restaurants, repair shops etc.

2)
In some surveys and in many censuses, a listing is first made by the enumerators. This requires:
- Accurate maps
- Good listing forms
- Good boundaries of the area to be listed
- Point to start listing and the listing procedure
- Decision how to list multiple dwellings in a building (basement, side entrances, door bells)

3)
In some surveys, the enumerator must select the units for enumeration. This is often the case in villages (secondary sampling units). The enumerator must then know how to select (randomly).

4)
In some surveys, the enumerator must make measurements. He must be able to count, make estimations, give personal judgments.

5)
In many surveys, the enumerator should do interviews with people. In this case, the enumerator should:
- show due respect for the respondent
- not show any official power
- show credential identity card
- explain the objective in the language which the respondent understands
- indicate that the information is confidential
- create a favorable atmosphere
- be properly dressed
- conform to local customs (remove shoes, evening interviews)

6)
The enumerator should also be able to check the completed forms before bringing them to the office.

SUPERVISORS

It is advisable to form small groups with a trained supervisor who is familiar with the survey. The supervisor should:
- anticipate problems before they arise
- control the enumerators
In the field, the accuracy of measurement, reliability and adherence to the selected sample should be controlled. Control should take place frequently in the beginning and later reduced.
- in some cases be used for call backs to hostile respondents.
- act as communication channel between administration and field workers.

ADMINISTRATION OF FIELD WORK

The administration should:
- Have a complete list of the names of interviewers and the addresses they will visit that day.
- Inform local authorities and police about the survey.
- Take care of the public relations. The more people know that there will be a survey, the higher will be, probably, the response rate.

PROBLEMS IN FIELD WORK

In field work you can meet the following problems:
- Failure to complete the interview.
- Resignation of enumerators during field work.
- Problems related to insufficient instruction of enumerator.
- Non-response.

All of these problems should be foreseen by good planning.

HOW TO DEAL WITH NON-RESPONSE

Non-response is a problem no investigator of human populations can escape. His survey material is not entirely under his control and he can never get information about more than a part from it.

We cannot simply ignore non-response, because the missing part is often different from the rest. If you perform a survey to find out how often people go to the cinema, than it goes without saying that people who are going to cinema every day are on average less at home than people not going at all to the cinema.

There are 6 types of non response:
- Difficult to reach (tribal area, war situation)
- Unsuitable for interview (illiterate, deaf)
- People who moved to another house
- Refusals
- People away from home during the whole survey
- People away from home only temporarily

30% is a very common non-response rate

There is not just one solution to non response. Of course, you may and should return to people who were not at home and you can substitute other people for the movers. However, it is most important to keep the questionnaire as brief as possible, to use professional enumerators and to provide even incentives to cooperate.

Assignment:
1) Collect data from at least 5 units of investigation with the help of the questionnaire for the Islamabad survey.
2) Record your response rate and write down any problems you encountered while surveying.

Lecture 6: Preparation of data

This lecture deals with:
- Planning of the processing
- Editing and coding of data
- Introduction of data in computer
- Security aspects

PLANNING OF THE PROCESSING

Processing of data is the transformation of data into usable information. Of course, the information is only usable when it arrives in time. A delay in the processing can be caused if:
- Data gathered do not meet the requirements.
- There are difficulties with the data processing techniques.
- There are not enough resources.

Planning can be a solution to these problems

Before the data processing can start, the following must be planned:
1) The output, should it be descriptive or analytical?
   If descriptive, frequency distributions can be used.
   If analytical, advanced statistical methods can be used.
2) The way to produce this output, which computer package to use etc., depends on quantity of data and output required.
3) How to record the data in order to facilitate processing?
   Should it be numeric, integer or text?
   Should one or more datafiles be used?
   Enables it the link with other surveys?
4) What are the resources to be used?
   Time, computer time and manpower must be planned.

Output required determines software and recording way. Software determines recording way.

EDITING AND CODING OF DATA

e d i t i n g

Editing is the detection and eventual correction of faulty data. It is one of the more boring parts of the survey process. Depending on the place and type of survey, wrong data should be send back to the enumerator or, alternatively, deleted.

Faulty data can be detected by too much homogeneity, too many zeros. Data should be checked on:

- Completeness

Is there an answer to each question?
If not, try to deduce from other questions what the answer should be. Maybe the interviewer still remembers the answer. If

too many questions are not available, the questionnaire should be regarded as lost.

The question may not have been filled in for the following reasons:
* Refusal of respondent to answer.
* Forgotten by enumerator.
* The question was not applicable to respondent.

- Accuracy
Are the data reliable? Inconsistencies between the answers can be found by comparing questions. This comparison can be done by means of computer. One of the programs which allow for this is Blaise (used by FBS). An example of an inconsistency is for example if a young person with a simple profession reports a very high income.

- Uniformity
Did each enumerator interpret the question in the same way? After the editing of the data a judgement must be made concerning the quality of the data and the fulfillment of the sample requirements. (sample size still sufficient, strata large enough). If the judgement is negative, it may be necessary to redo part of the survey.

c o d i n g

A very simple example of coding is the following: Male = 1, female = 0. The numbers, rather than the words, will be introduced in the computer. The advantage is that it will be faster to introduce the data and that computer space will be saved.

For each questionnaire there is a corresponding coding frame. The coding frame gives the explanation of the codes used. It should be in accordance with other surveys if possible. It should start with an unique identification code of the questionnaire. The frame can relate to a single question, groups of questions or the entire questionnaire. The distance between the filled in questionnaire and the coded data can be made smaller by the simultaneous design of questionnaire and coding frame.

Some of the coding can be done by enumerators, the remaining part by the people entering the data in the computer. Sometimes, it may be necessary to introduce a separate stage for coding and only then introduce it in the computer. The coding can be done on the questionnaire itself or on special forms.

The coding of open questions is more difficult than that of closed questions because it must allow for the different dimensions of answers.

Look at the following open question:

In what way you are affected by the rising prices?
The answers can have the following dimensions:
- general costs of living increase, it is more difficult to meet the ends.
- costs of production are increasing (for producer).
- change in purchasing habits (I do not buy this or that anymore) etc.

INTRODUCTION OF DATA IN COMPUTER

Once the codes are easy to read from the questionnaire or specially designed coding forms, we can introduce the data in the computer. Obviously, how this is done depends on the type of soft-ware chosen. In this course we will use Reflex2 to introduce and analyse data.

SECURITY ASPECTS

The security aspects in data processing are very important. Frequent back-ups must be made. Only authorized personnel should have access to the data. Confidentiality must be guaranteed. Storage of data, also after the project, must be arranged well.

Assignments:
1) Write down the explanation of codes used in the questionnaire.
2) Introduce the data in the computer.
3) Edit the data.
4) Make a back up of the data on floppy disk.

Lecture 7: Data processing

This lecture deals with:
- Analysis of data

ANALYSIS OF DATA

The analysis of data is dependent on the type of survey and the type of variables gathered. Advanced statistical methods can often only be used in a reliable way for larger data sets.

In many survey reports analysis does not go beyond distributions, percentages, averages and measures of dispersion, supplemented by suitable diagrams. Of course, in such reports breakdowns are used. Very often, questions are cross-classified by age, sex and social class. In Labour statistics often by occupation, education, city etc.

Interpretation and analysis should be preferably done by the surveyor himself, because he knows the background of the survey. Interpretation and analysis is a very skillful task.

Analysis of data includes:
- inferences from sample to population
- interpretation of relationships
- Some advanced statistical methods for testing significance
- multivariate analysis

Inference from sample to population can create substantial problems if difficult designs with several stages, phases, stratifications and varying probabilities of selection are used. However, even with EPSEM sampling (equal probability of selection for each member) there are many hazards which may arise. The loss of coverage due to incomplete sample frames or non-response must be acknowledged and the conclusions generalized only to the population actually covered.

Similarly, when a survey has for reasons of convenience or lack of resources been confined to, say, one town, the researcher must resist the temptation to generalize the conclusions to the entire country. He must resist even when the town seems very typical on various indicators.

Example: Mortality of left handed people

According to conclusions from a survey performed by two American researchers, left handed people life on average nine years less than right handed people.
On what did they base this conclusion?

1) 2,875 mail questionnaires to families of people who just died.However, only 36% (1,033) were returned. Hence, a high non-response.
2) 7 to 8% was left handed. Hence, only 75 to 80 persons were left handed. This may be statistically significant, but is a small sample.
3) The sample was taken from two cities in California only. As we do not know what factors have an impact on left handedness, it may be very dangerous to extend this to the entire humanity.
4) American population statistics show that share of left handed people decrease with increasing age. Netherlands population statistics show a similar, but much less rapidly decreasing share of left handed people with growing age. This can be explained by social pressure in the past not to be left handed. In China, social pressure proved to be very effective.

Even if the figures are true, left handed participants do not have to be worried. Netherlands figures show that the percentage left handed people with university education is 22% among older people as compared to 11% for the rest of the population. So, left handed intellectuals live longer.

Interpretation of relationships

Even if there is a strong association between two variables, this does not automatically mean that there is also a cause-effect relation between them. As a statistician, you often have to be extremely careful.

For example, let us take the connection between smoking and lung cancer. Various studies have shown a positive association. However, they do not show whether lung cancer is causing people to smoke or that smoking is causing lung cancer. More important, it does not show whether there is maybe a third variable having an impact on lung cancer and smoking. For example, nervous people may both smoke more and have a higher chance of getting lung cancer. In this case nerves rather than smoking is causing cancer.

Another causal relation may be that smoking makes people nervous which is causing cancer. Again nerves rather than smoke cause cancer.

Some advanced statistical methods for testing significance

See reader on statistical methods for sample surveys

multivariate analysis

See reader on statistical methods for sample surveys

Assignment:
1) Process the data with the help of the computer program Reflex. Make the necessary cross tabulations.

Advanced assignment:
2) Use the statistical manual to calculate whether the means in two different strata for the same variable differ significantly from each other on the 95% level.

Suggestions for reading
Moser and Kalton, Survey Methods in Social Investigation, ch 17

Stopher, P. and A. Meyburg, Survey Sampling and Multivariate Analysis for Social Scientists and Engineers. D.C.Heath, Lexi-ngton, 1979.

Lecture 8: Reporting

This lecture deals with:
- Rules for report writing
- Structure of reports

RULES FOR REPORT WRITING

The most important output of the survey is the systematic presentation of its main data and conclusions in a written form.

Systematic means that the report has a logical structure and deals with separate research issues in separate chapters. The report must be concise and clear.

Written means that not only tables are included in the report, but that they are explained by means of supplementary text as well. The reader should be helped to read and analyse the tables in a very short time.

Of course, the report must answer the demands raised by the customer, (policy maker, firm etc.).

Especially in the case of very large surveys, often different types of reports are written: preliminary, technical, and general reports. Most of below mentioned remarks are valid here, but the structure and language in these reports may deviate (see UN recommendations for report writing).

STRUCTURE OF REPORTS

In the report, there should be at least:

- Front page
  including title, authors, issuing institution, date.

- Table of contents
  with the heading Contents (in lower case letters, underlined). It contains the numbers (followed by a dot) and titles of the sections and subsections. Next, the words references and appendices may follow. Titles of subsections are indented. Page numbers refer to the respective places.

- Summary and Conclusions
  After the table of contents often the summary of the survey is presented, including:
  - Where and in what context the survey was carried out.
  - The purposes of the survey.
  - The contents of the chapters in the paper with the main results per characteristic.

  Summary and conclusions can also be presented at the end of the report.

- Introduction
  In the introduction of the paper, the purposes and context of the survey are described more extensively. The main problem that was to be investigated by the survey should receive due attention. It is not wrong to repeat the purpose of the survey both in summary and introduction.

- Abstract of questionnaire

- First section, justification of survey methods

  The design of the survey should follow from the purpose of the survey. Sample size, frame and design used should be justified. The coverage and representativeness of the survey should receive attention. All known aggregate data concerning the population from which the sample was drawn must be given (for example total number of firms with their total employment). The stratification used must be justified and a table with units of investigation per strata can be given (for example firms by activity).

- Other sections, results

  The discussion of survey results includes:
  - Selection of results: the survey always contains more data than the report can include. Selection of data should primarily take the purpose of the survey into account.
  - Presentation of results includes the grouping of data in suitable tables.
  These tables have been planned before and may be adjusted. The main table must be in the first analyzing section. Other chapters should discuss relevant subdivisions, always in relation to the original problem.

  Standard errors can best be presented by translating into confidence intervals. Is the result significant on 95% or 99% level?

- Interpretation of results (See survey analysis.)

Assignment: Write a small report on the base of the prepared tables. Do not forget to include a small introduction (describing purpose of survey, concepts and definitions and survey methods), conclusions and the full questionnaire