# Figure 3.1b. SAMPLE SURVEY DESIGN/EXECUTION: Introductory Illustrations

The excerpt reprinted in this Figure 3.1b is from *Practical Statistics Simply Explained: The Principles and Practice of Statistical Inference* by Russell Langley, Dover Scientific Books, 1971, pages 102-108, a later version of Pan Books, 1968 and 1970.

**Sample Surveys**

There are three types of sample surveys.

- Retrospective
- Current
- Prospective

*Retrospective surveys* look back over information recorded in the past. The usefulness of this type of investigation is often limited by records being incomplete in respect of the data being sought. It is hard to prevent this fault, for who knows what data may interest us in 10 years' time, and it is costly to keep records of things that will never be wanted. Nevertheless, good work has been done with this technique, as for instance:

During the Second World War, it was necessary to keep planes in action as much as possible, so it was decided to see if the number of time-consuming engine overhauls could be reduced without risk. A retrospective survey was made of planes that were lost and, contrary to all expectations, it was found that the number of planes lost as a result of engine troubles was greatest right after overhaul, and actually decreased as the time since overhaul grew longer. This result led to a considerable increase in the intervals between overhauls and, needless to say, to important revisions in the manner of overhauling to make sure that all those nuts and bolts were tightened up properly. (from W.A. Wallis and H.V. Roberts, *Statistics – A New Approach*, Free Press, 1960).

A more extended example is discussed in Wallis and Roberts' book §2.8.2. It concerns a retrospective survey of the incidence of insanity a century ago, to compare with the present rate, as part of a project to assess the influences of the pressures of modern living on mental health. A sound statistical approach has elevated this particular survey into a most important sociological document.

*Current surveys* are those which seek information about things as they are at the present time. The two surveys of radio audiences described on page 100 are typical examples.

*Prosepctive surveys* are a fairly new idea. They compare information obtained about some subject now, with a follow-up about the same subject obtained from the same individuals at a later date. The American Cancer Society's investigation described on page 37 is an example of a prospective survey.

Prospective and current survys are beset with similar problems, so they will be discussed together. These problems concern the design of the questionnaire, the sampling, and the actual conduct of the survey.

Having decided exactly what is being sought from the survey, the next thing is to prepare a question or, more usuallly, a set of questions, designd to bring forth the desired information. The questions must be worded in simple, clear language. Standard answers should be provided whenever possible ('circle the answer which applies in your case'), otherwise you may find that in answer to a question such as: 'Where do you get chillblains?,' one person may reply 'Lower limbs', another may reply, 'On left big toe, ¼ inch from cuticle', while yet aother may answer, 'Only at ski resorts'. After wording all the questions as carefully as possible, the next step is to try out the questionnaire on a number of people, for the express purpose of detecting unsuspected ambiguities or misunderstandings. Some people would be unsure about this question: 'What medications, if any, do you take regularly?' They may be unaware as to whether 'medicine' refers only to a medicine prescribed by their doctor or whether it also covers some mild patent remedy, such as throat lozengers or a laxative. Again, some questions to which a simple yes or no answer is expected may give trouble if the person finds that to be correct he should answer it 'sometimes yes, sometimes no'. An example of such a question would be, 'Do you usually take notes in lectures?' Perhaps the answer will depend on who the lecturer is or whether the subject is satisfactorily dealt with in textbooks, rather than on the habits of the person replying to the questionnaire. Maybe it would be better to change the question to read, 'Do you take lecture notes (a) always, (b) usually, (c) sometimes, (d) rarely, or (e) never?'

After this pre-testing, the questionnaire can be revised and, if necessary, retested to make sure it is faultless.

Even so, some of the answers will be wrong. Ladies will understate their age, gentlemen will overstate their occupation (the man who repairs electric toasters may elevate his status to 'electrical engineer'), and some adults might even be reluctant to admit that the main reason they buy a certain newspaper is because they like its comic strips.

Vance Packard in *The Hidden Persuaders* (Penguin, 1962) quotes some nice cases. For example, a survey was carried out on behalf of a firm that sold kippered herrings, in order to discover why sales were lagging. It turned out that most of the people interviewed said they just didn't like the taste of kippers. However, under persistent and subtle probing, it was discovered that 40% of these people who said they didn't like kippers had never tasted kippers in their lives!

A manufacturer made two sizes of kitchen-range, one large, the other small. The large model proved much the better seller of the two, and a questionnaire revealed that housewives preferred it because it had much more working space on its top. Taking a tip from this, the firm then designed a medium-sized, lower-priced stove with an exceptionally large working space. Imaginge their disappointment when they found that this new stove just wouldn't sell. The manufacturer then consulted a market research firm to investigate the mattter. The research firm's psychologist found that the housewives were not consciously telling fibs (as in the case of the kippered herrings, above), but the story about the large working area on the stove top was really only a rationalization of their unconscious desire for a big expensive-looking stove. [These excerpts are reprinted by permission of the copyright holders, Longmans, Green & Co. Ltd., London, and David McKay Co. Inc., New York.]

Let these examples be a lesson to anyone naive enough to believe that people always tell the truth to opinion interviewers or questionnaires. In market research, these little 'fibs' can be very costly to the misled manufacturer, so psychologists are often brought into the affair. Their questions probe beneath the surface. Once they used to ask, 'We are thinking of marketing a new talc powder in a red tin, and would appreciate your opinion; do you think red would be a nice colour for the tin?' Nowadays a psychologist might ask exactly the same question in the following form: 'Do you associate the colour red with any particular part of the human body?'

To illstrate the potential value of depth probing, Vance Packard quotes the case of a psychologist who gave a Rorschach ink-blot test to 80 smokers who had a strong loyalty to one or other brand of cigarettes, and from this test was able to name their favourite brands in nearly every single case. The ink-blot test shows certain personality traits, which he was able to link with the various brand 'images' of the cigarettes.

However, most current and prospective surveys don't need any depth probes, just plain straight questions and answers.

Sampling is the next problem. As explained on page 41, the selecting of a proper sample, especially of people, is not easy. A stratified random sample is generally the best kind for surveys, particularly as the sample size required for a given degree of precision is quite a lot smaller than would be the case with an ordinary random sample (page 40). But it's easier to talk about getting a good sample than it is to get one. It involves a continuous battle against factors which upset randomness. But the sample must be random.

Non-response is one of the main factors



(*continued overleaf*)

which will ruin the randomness of any sample, no matter how well-chosen the sample may be. All those people who are not at home, who would rather not answer that particular question, who are too sick to answer any questions, or who don't know the answers to some of the questions, all these people had better be few and far between, or your sample could become seriously biased. Because the risk is always there that these non-respondents may differ significantly from those who do respond. Hence, this fault cannot be rectified by simply questioning more people to allow for this 'shrinkage'.

The first Kinsey Report, a survey of the sexual habits of over 12,000 American men, published in 1948, came under considerable criticism for ignoring the problem of non-response. The authors of the Report say little about the method used to select their sample, except to admit that randomness is hard to achieve. But we know that the mere size of their sample does not make it a random one (page 38), so we are left to wonder whether those men who excluded themselves from the sample, *i.e.*, the non-respondents, would have changed the sample averages or not.

The problem of non-response should never be ignored. It should be tackled by spending more time and money to reduce this pool of resistance to an absolute minimum; this may mean that the sample will be a smaller one, but the data obtained will be better.

The sample size needed depends on what is being sought and on the degree of precision required. If the sample is random, its reliability increases with the square root of the number in the sample; if the sample is not random, its reliability remains a matter of luck, regardless of the sample size. We have discussed this matter on page 45. To save wasting time and money, a statistician should be consulted to make an estimate of the sample size; otherwise, the survey may turn out to be too small for any reliance to be placed on its answers, or it may end up unnecessarily large.

Wallis and Roberts recall a statistical traffic jam which occurred in New York State in 1950. In order to find out the most suitable route for a new highway between New York City and Buffalo, a traffic commission decided to conduct a 24-hour survey of vehicles using the existing road. So, on a certain Sunday, every fourth vehicle was stopped and the driver was asked three questions:

(1) Where did you come from?
(2) Where are you going?
(3) How often do you make this trip?

It seemed simple enough. but the objective was largely defeated by acute oversampling. It caused a traffic jam 10 miles long and many motorists, hearing of the jam, no doubt took alternative routes. The 25% sample also was much larger than necessary. It would have been a better design to have taken a smaller survey, say every 100th vehicle, and spread the investigation over a whole week.

Despite a goodly number of formulae to help them estimate the sample size, statisticians will often ask for a small preliminary survey to be conducted; the results of this are then used to increase the accuracy of their estimation.

The actual conducting of the survey is generally done either by personal interviews or by mail. Mailing the questionnaire is cheaper, but the percentage of non-responses may be high. It may take several letters, a telegram, or phone call, and finally sending an interviewer to get all the answers back. The *Literary Digest* didn't do these things (p. 38).

The efficacy of different interviewers varies quite a lot. Some have the knack of putting people at ease and getting the answers quickly and easily; others haven't. Sometimes, the interviewer unwittingly affects the answers; for exmple:

Some years ago, the National Opinion Research Centre in America found that when African Americans were asked, 'Do you think that African Americans are getting a fair deal in the Army?', 35% said No to African American interviewers while only 11% said No to white interviewers.

Of course, interviewers must be properly trained. For instance, they must never ask leading questions (*i.e.*, questions worded in such a way that the answer is suggested) such as, 'After work, do you prefer fancy drinks like cocktails, or a man-sized glass of beer?' It'd take a tough customer to admit to a cocktail under these conditions, wouldn't it? Or how about the following more subtle example:

*Interviewer*: What brand of toothpaste do you use?
*Lady*: Densol.
*Interviewer*: Why do you like Densol?
*Lady*: Ah...um...I've never really thought about it.
*Interviewer*: Well, try to give me an answer. (Pause) Is it because of its taste?
*Lady*: Yes, it's got a nice taste.

And the interviewer writes down, 'Uses Densol because she likes the taste'. But perhaps the lady wasn't prepared to admit the real reason, that it was the cheapest brand, or that she thought it counteracted her bad breath.

Any way you look at it, even a well-trained interviewer is still a human factor situated at a key point in all personal surveys.

These, then, are the kinds of things that must be dealt with in designing surveys.

1 Comment on the *statistical* issues involved in an investigation (like that of Wallis and Roberts, mentioned in the third paragraph of the left-hand column overleaf on page 3.5) of possible changes over time in the prevalence, and the causes of, insanity.

2 In the language of the FDEAC cycle, rephrase the first statement starting at the bottom of the left-hand column overleaf on page 3.5: *Having decided exactly what is being sought from the survey, ...*

3 Comment critically on the last word 'faultless' in the second (short) paragraph of the middle column overleaf on page 3.5.

4 Starting in the second-last paragraph of the right-hand column overleaf on page 3.5, and continuing above, the article emphasizes the importance of *random* selecting [what *we* call equiprobable selecting (EPS)].

- Whar reason is given in the article for this importance?
- In our terminology, which category of error is involved?
- Explain briefly the *statistical* issues involved in the statement (and its wording) starting near the bottom of the left-hand column above: *if the sample is not random, its reliability remains a matter of luck, regardless of the sample size.*

5 Distinguish clearly our two categories of error involved in the statement starting in the last paragraph overleaf on page 3.5: *Non-response is one of the main factors which will ruin the randomness of any sample, no matter how well-chosen the .....*

6 Explain briefly the trade-off in statistical issues involved in the final statement in the third paragraph of the left-hand column above: *this may mean that the sample will be a smaller one, but the data obtained will be better.*

7 Explain briefly the meaning of the final statement in the third paragraph of the middle column above: *the results of this (pilot survey) are then used to increase the accuracy of their estimation.*