

# **IMPORTANCE OF DATA IN RESEARCH**

## **The importance of good quality data**

Scientific research is used by academics in a wide scope of academic disciplines such as social sciences, public health, biostatistics, education, social work, public administration, and business administration; and by practitioners engaged in marketing, commerce, and industry. Data are the basis for all scientific research. Collecting good quality data plays a vital role in supplying objective information for the problems under study so that some analytical understanding of the problems and hence solutions can be obtained. Making decision on the basis of poor quality data is risky and may lead to disastrous results, as the situation may be distorted and hence all subsequent analyses and decision making will rest on a shaky ground.

## **Statistical sampling techniques**

Sampling plays a vital role in data collection. Drawing a representative sample saves time, money and efforts in research and helps achieve desirable degree of reliability on collected data. There are many types of sample design and choosing the appropriate design can improve the accuracy and reliability of sample statistics while keeping the cost of survey low.

## **Designing a good questionnaires**

A lot of expertise is required in designing a good questionnaire. Factual questions requiring memory may produce answers with big memory errors. Opinion questions improperly asked may lead to biased results. Questions with unspecific, complex, ambiguous wordings may produce poor quality data. Question order also may affect the answers obtained, especially when one is concerned with opinions that are unstable.

## **Conducting surveys**

To conduct a survey, the aims must be clearly stated. The accuracy required, and the cost, time and manpower available need to be estimated. The population to be studied should be clearly defined. Whether personal interview, telephone interview or mailed questionnaire is to be used needs to be considered. A good questionnaire is to be designed. The likely sources of error of the survey data and the precautions that can be taken to minimized them should be considered. A

pilot test should be performed so that the survey procedures can be finalized. Before the fieldwork is conducted, training may be given to interviewers to guarantee their quality. The method of field supervision has to be decided. Methods for processing and analyzing the survey data should also be considered.

**Source:** Statistical Advisory Unit, Department of Applied Mathematics, The Hong Kong Polytechnic University

### **Validity, reliability and accuracy**

Social science research confers a special meaning to validity:

'the extent to which a measure, indicator or method of data collection possesses the quality of being sound or true as far as can be judged. ... in the social sciences generally, the relationship between indicators and measures and the underlying concepts they are taken to measure is often contested' (Jary & Jary, 1995: 714).

In effect, the validity of information is its relevance and appropriateness to your research question and the directness and strength of its association with the concepts under scrutiny. Often you will have to use best available information whose validity may be weak. For example, to what extent, if any, does the decline in 'sectarian violence' in N. Ireland post-2001 reflect a lessening of antagonisms between conflicting groups? Does the election of an opposition party reflect popular support for its manifesto or criticism of the outgoing government? Do declining rates of party membership reflect a lessening of interest in health and education? One measure that intrigues Politics researchers is the counterfactual – events that don't happen – as evidence of hegemonic domination.

But how can researchers be confident that the absence of an event can be attributed to the omnipresence of another? One solution to this particular problem of problematic validity is for you to adopt a wider range of measures to reduce dependence on any one.

**Reliability** is, literally, the extent to which we can rely on the source of the data and, therefore, the data itself. Reliable data is dependable, trustworthy, unfailing, sure, authentic, genuine, reputable.

Consistency is the main measure of reliability. So, in literary accounts, the reputation of the source is critical.

In John Cole's view, Richard Crossman was not a reliable diarist. Indicators of reliability will include proximity to events, (whether the writer was a participant or observer,) likely impartiality, and whether, as the police say, the record was really contemporaneous or an eventide reflection on the day's events. Very

few politicians admit to real failings: all too often, their own agenda appears to justify their actions or to criticise others.

Accuracy is sensitivity to change – especially of detail, e.g. dates, numbers, persons present, etc. Remember that some biographers deliberately add false detailed information to trap and sue plagiarisers

### **Facts and truth**

Once again, you will find that adopting a critical distinction between facts and truth is useful.

Facts are the available data. They present incomplete snapshots of events. Truth is the reality behind the facts. Sometimes the facts may obscure the truth –perhaps deliberately so. A good example was provided to me by a leading academic. He privately described how he had critically reviewed a best-selling account of British rural life where that the author had misrepresented the facts by combining material from a number of interviews to represent a composite figure. The author had replied to the effect that his critic was unable to distinguish between the facts and truth