Sources acquisition and classification of Data

ugcnetpaper1.com/sources-acquisition-and-classification-of-data/

July 3, 2020

Sources, acquisition and classification of Data

Unit-VII Data Interpretation

Data Sources & acquisition

Data can be defined as the quantitative or qualitative values of a variable. Data is plural of datum which literally means to give or something given. Data is thought to be the lowest unit of information from which other measurements and analysis can be done.

Data can be numbers, images, words, figures, facts or ideas. Data in itself cannot be understood and to get information from the data one must interpret it into meaningful information. There are various methods of interpreting data.

Data sources are broadly classified into primary and secondary data.

Data collection plays a very crucial role in the statistical analysis. In research, there are different methods used to gather information, all of which fall into two categories, i.e. primary and secondary data (Douglas, 2015).

As the name suggests, primary data is one which is collected for the first time by the researcher while secondary data is the data already collected or produced by others.

ugcnetpaper1.com/sources-acquisition-and-classification-of-data/

1/6

There are many differences between primary and secondary data, which are discussed in this work. But the most important difference is that—

- primary data is factual and original whereas secondary data is just the analysis and interpretation of the primary data.
- While primary data is collected with an aim for getting a solution to the problem at hand, secondary data is collected for other purposes.
- The fundamental differences between primary and secondary data are; the term primary data refers to the data originated by the researcher for the first time while secondary data is the already existing data collected by the investigator agencies and organizations earlier.
- Primary data sources include surveys, observations, experiments, questionnaire, personal interview etc. on the other contrary, secondary data collection sources are government publications, websites, books, journal articles, internal records etc.

PRIMARY DATA

Primary data means original data that has been collected specially for the purpose in mind. It means someone collected the data from the original source first hand.

Data collected this way is called primary data. Primary data has not been published yet and is more reliable, authentic and objective. Primary data has not been changed or altered by human beings; therefore, its validity is greater than secondary data.

SECONDARY DATA

Secondary data is the data that has been already collected by and readily available from other sources. When we use Statistical Method with Primary Data from another purpose for our purpose we refer to it as Secondary Data. It means that one purpose Primary Data is another purpose Secondary Data. So that secondary data is data that is being reused. Such data are more quickly obtainable than the primary data.

These secondary data may be obtained from many sources, including literature, industry surveys, compilations from computerized databases and information systems, and computerized or mathematical models of environmental processes.

Data acquisition

Data is one of the most important and vital aspects of any research studies. Researchers conducted in different fields of study can be different in methodology but every research is based on data which is analyzed and interpreted to get information. Data is the basic unit in statistical studies. Statistical information like census, population variables, health statistics, and road accidents records are all developed from data.

There are two sources of data collection techniques. Primary and Secondary data collection techniques, Primary data collection uses surveys, experiments or direct observations.

Secondary data collection may be conducted by collecting information from a diverse source of documents or electronically stored information, census and market studies are examples of common sources of secondary data. This is also referred to as "data mining."

Survey-

The survey is the most commonly used method in social sciences, management, marketing and psychology to some extent. Surveys can be conducted in different methods.

Questionnaire-

The questionnaire is the most commonly used method in the survey. Questionnaires are a list of questions either an open-ended or close-ended for which the respondent give answers. A questionnaire can be conducted via telephone, mail, live in a public area, or in an institute, through electronic mail or through fax and other methods.

Interview-

An interview is a face-to-face conversation with the respondent. It is slow, expensive, and they take people away from their regular jobs, but they allow in-depth questioning and follow-up questions.

Observations–

Observations can be done while letting the observing person know that he is being observed or without letting him know. Observations can also be made in natural settings as well as in the artificially created environment.

Published Printed Sources

Related Posts
[Solving Tricks | Data Interpretation NET EXAM | Updated 2020](#)

[Data and Governance | New Topic in UGC NET EXAM](#)

There are varieties of published printed sources. Their credibility depends on many factors. For example, on the writer, publishing company and time and date when published. New sources are preferred and old sources should be avoided as new technology and researches bring new facts into light.

Books

Books are available today on any topic that you want to research. The uses of books start before even you have selected the topic. After selection of topics books provide insight on how much work has already been done on the same topic and you can prepare your literature review. Books are a secondary source but most authentic one in secondary sources.

Journals/periodicals

Journals and periodicals are becoming more important as far as data collection is concerned. The reason is that journals provide up-to-date information which at times books cannot and secondly, journals can give information on the very specific topic on which you are researching rather talking about more general topics.

Magazines/Newspapers

Magazines are also effective but not very reliable. Newspaper, on the other hand, is more reliable and in some cases, the information can only be obtained from newspapers as in the case of some political studies.

Published Electronic Sources

As internet is becoming more advance, fast and reachable to the masses; it has been seen that much information that is not available in printed form is available on internet. In the past the credibility of internet was questionable but today it is not.

The reasons that in the past journals and books were seldom published on internet but today almost every journal and book is available online. Some are free and for others you have to pay the price.

E-journals: e-journals are more commonly available than printed journals. Latest journals are difficult to retrieve without a subscription but if your university has an e-library you can view any journal, print it and those that are not available you can make an order for them.

General Websites; Generally, websites do not contain very reliable information so their content should be checked for the reliability before quoting from them.

Weblogs: Weblogs are also becoming common. They are actually diaries written by different people. These diaries are as reliable to use as personal written diaries.

Classification of Data

Data classification is the process of organizing data into categories for its most effective and efficient use.

Classification is the way of arranging the data in different classes in order to give a definite form and a coherent structure to the data collected, facilitating their use in the most systematic and effective manner. It is the process of grouping the statistical data

under various understandable homogeneous groups for the purpose of convenient interpretation.

There are three different approaches are the industry standard for data classification:

- Content-based classification
- Context-based classification
- User-based classification

Objectives of classification of data:

- To group heterogeneous data under the homogeneous group of common characteristics;
- To facility similarity of the various group;
- To facilitate effective comparison;
- To present complex, haphazard and scattered dates in a concise, logical, homogeneous, and intelligible form;
- To maintain clarity and simplicity of complex data;
- To identify independent and dependent variables and establish their relationship;
- To establish a cohesive nature for the diverse data for effective and logical analysis;
- To make logical and effective quantification
- A good classification should have the characteristics of clarity, homogeneity, and equality of scale, purposefulness, accuracy, stability, flexibility, and unambiguity.

Classification is of two types, viz., quantitative classification, which is on the basis of variables or quantity; and qualitative classification (classification according to attributes). The former is the way of grouping the variables, say quantifying the variables in cohesive groups, while the latter group the data on the basis of attributes or qualities. Again, it may be multiple classification or dichotomous classification.

The former is the way of making many (more than two) groups on the basis of some quality or attributes, while the latter is the classification into two groups on the basis of the presence or absence of a certain quality.

Data classification, in the context of information security, is the classification of data based on its level of sensitivity and the impact to the University should that data be disclosed, altered or destroyed without authorization. The classification of data helps determine what baseline security controls are appropriate for safeguarding that data. All institutional data should be classified into one of three sensitivity levels, or classifications:

A. Restricted Data

Data should be classified as Restricted when the unauthorized disclosure, alteration or destruction of that data could cause a significant level of risk to the University or its affiliates. Examples of Restricted data include data protected by state or federal privacy regulations and data protected by confidentiality agreements. The highest level of security controls should be applied to Restricted data.

B. Private Data

Data should be classified as Private when the unauthorized disclosure, alteration or destruction of that data could result in a moderate level of risk to the University or its affiliates. By default, all Institutional Data that is not explicitly classified as Restricted or Public data should be treated as Private data. A reasonable level of security controls should be applied to Private data.

C. Public Data

Data should be classified as Public when the unauthorized disclosure, alteration or destruction of that data would results in little or no risk to the University and its affiliates. Examples of Public data include press releases, course information and research publications. While little or no controls are required to protect the confidentiality of Public data, some level of control is required to prevent unauthorized modification or destruction of Public data.

Quantitative and Qualitative Data

- Quantitative data are anything that can be expressed as a number, or quantified. Examples of quantitative data are scores on achievement tests, the number of hours of study, or weight of a subject. These data may be represented by ordinal, interval or ratio scales and lend themselves to most statistical manipulation.
- Qualitative data cannot be expressed as a number. Data that represent nominal scales such as gender, social economic status, religious preference are usually considered to be qualitative data.

Reference –

- https://www.slideshare.net/manukumarkm/source-of-data-in-research
- Data Classification- https://www.cmu.edu/iso/governance/guidelines/data-classification.html