

PROCESSING OF DATA, EDITING, CODING, & CLASSIFICATION

After the data have been collected, the researcher turns to the task of analyzing them. The data, after collection, has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis. Thus, a researcher should classify the raw data into some purposeful and usable categories. Coding operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted. Editing is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, especially in large inquiries, is tabulated by computers. Computers not only save time but also make it possible to study a large number of variables affecting a problem simultaneously. Analysis work after tabulation is generally based on the computation of various percentages, coefficients, etc., by applying various well defined methods or techniques. Supporting or conflicting with original new hypotheses should be subjected to tests of significance to determine with what validity data can be said to indicate any conclusion(s).

DATA ANALYSIS BASICS: EDITING, CODING, & CLASSIFICATION

After collecting data, it must be reduced to some form suitable for analysis so that conclusions or findings can be reported to target population. For analyzing data researchers must decide –

- (a) Whether the tabulation of data will be performed by hand or by computer.
- (b) How information can be converted into a form that will allow it to be processed efficiently.
- (c) What statistical tools or methods will be employed.

Now a days computers have become an essential tool for the tabulation and analysis of data. Even in simple statistical procedures computer tabulation is encouraged for easy and flexible handling of data. Micro and laptop computers can produce tables of any dimension and perform statistical operations much more easily and usually with far less error than is possible manually. If the data is large and the processing undertaken by computer the following issues are considered.

1. Data preparation which includes editing, coding, and data entry.
2. Exploring, displaying and examining data which involves breaking down, examining and rearranging data so as to search for meaningful description, patterns and relationships.

1.EDITING

First step in analysis is to edit the raw data. Editing detects errors and omissions, corrects them whatever possible. Editor's responsibility is to guarantee that data are – accurate; consistent with the intent of the questionnaire; uniformly entered; complete; and arranged to simplify coding and tabulation. Editing of data may be accomplished in two ways - (i) field editing and (ii) in-house also called central editing. Field editing is preliminary editing of data by a field supervisor on the same data as the interview. Its purpose is to identify technical omissions, check legibility, and clarify responses that are logically and conceptually inconsistent. When gaps are present from interviews, a call-back should be made rather than guessing what the respondent would probably said. Supervisor is to re-interview a few respondents at least on some pre-selected questions as a validity check. In center or in-house editing all the questionnaires undergo thorough editing. It is a rigorous job performed by central office staff.

2.CODING: Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristic of exhaustiveness (i.e., there must be a class for every data item) and also that of mutual exclusivity which means that a specific answer can be placed in one and only one cell in a given category set. Another rule to be observed is that of unidimensionality by which is meant that every class is defined in terms of only one concept. Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. This makes it possible to precode the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires. But in case of hand coding some standard method may be used. One such standard method is to code in the margin with a coloured pencil. The other method can be to transcribe the data from the questionnaire to a coding sheet. Whatever method is adopted, one should see that coding errors are altogether eliminated or reduced to the minimum level.

3. CLASSIFICATION: Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

(a) Classification according to attributes: As stated above, data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.). Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain attributes are known as statistics of attributes and their classification is said to be

classification according to attributes.

Such classification can be simple classification or manifold classification. In simple classification we consider only one attribute and divide the universe into two classes—one class consisting of items possessing the given attribute and the other class consisting of items which do not possess the given attribute. But in manifold classification we consider two or more attributes simultaneously, and divide that data into a number of classes (total number of classes of final order is given by 2^n , where n = number of attributes considered).* Whenever data are classified according to attributes, the researcher must see that the attributes are defined in such a manner that there is least possibility of any doubt/ambiguity concerning the said attributes.

(b) Classification according to class-intervals: Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomenon which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and are classified on the basis of class intervals.