

IEH5F2– Metodologi Penelitian


Processing and Analysis of Data

Dr. Ir. Luciana Andrawina, MT

Prodi S2 Teknik Industri – Fakultas Rekayasa Industri



1. Editing
2. Coding
3. Classification
4. Tabulation



Editing of data is a process of examining the collected raw data (specially in surveys) to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaires and/or schedules. Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as completed as possible and have been well arranged to facilitate coding and tabulation.

▶ 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.

▶ 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.

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.

Tabulation is the process of summarising raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis

- ▶ (a) *Classification according to attributes:*
- ▶ (b) *Classification according to class-intervals:*

▶ 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.

- ▶ 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.

- ▶ (i) How many classes should be there? What should be their magnitudes?
- (ii) How to choose class limits?
- (iii) How to determine the frequency of each class?

(i) How many classes should be there? What should be their magnitudes?

▶ Determining the size of class interval (H.A. Sturges)

$$i = R / (1 + 3.3 \log N)$$

where

i = size of class interval;

R = Range (i.e., difference between the values of the largest item and smallest item among the given items);

N = Number of items to be grouped.

(ii) How to choose class limits?

While choosing class limits, the researcher must take into consideration the criterion that the mid-point (generally worked out first by taking the sum of the upper limit and lower limit of a class and then divide this sum by 2) of a class-interval and the actual average of items of that class interval should remain as close to each other as possible.

Consistent with this, the class limits should be located at multiples of 2, 5, 10, 20, 100 and such other figures.

Exclusive type class intervals: They are usually stated as follows:

10–20

20–30

30–40

40–50

The above intervals should be read as under:

10 and under 20

20 and under 30

30 and under 40

40 and under 50

Inclusive type class intervals: They are usually stated as follows:

11–20

21–30

31–40

41–50

(iii) How to determine the frequency of each class?

An

<i>Income groups (Rupees)</i>	<i>Tally mark</i>	<i>Number of families or (Class frequency)</i>
Below 400	III III III	13
401-800	III III III III	20
801-1200	III III II	12
1201-1600	III III III III	18
1601 and above	III II	7
Total		70

- ▶ Tabulation is essential because of the following reasons.
 1. It conserves space and reduces explanatory and descriptive statement to a minimum.
 2. It facilitates the process of comparison.
 3. It facilitates the summation of items and the detection of errors and omissions.
 4. It provides a basis for various statistical computations.

IEH5F2 METODOLOGI PENELITIAN

MATERI

Processing and Analysis of Data

ELEMENTS/TYPES OF ANALYSIS

mean the computation of certain indices or measures along with searching for patterns of relationship that exist among the data groups. Analysis, particularly in case of survey or experimental data, involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences.

A cluster of red triangles of various sizes and orientations in the top-left corner of the slide.


Analysis may, therefore, be categorised as descriptive analysis and inferential analysis (Inferential analysis is often known as statistical analysis). “*Descriptive analysis* is largely the study of distributions of one variable.

this sort of analysis may be in respect of one variable (described as unidimensional analysis), or in respect of two variables (described as bivariate analysis) or in respect of more than two variables (described as multivariate analysis

A cluster of red triangles of various sizes and orientations in the bottom-right corner of the slide.

▶ the joint variation of two or more variables for determining the amount of correlation between two or more variables.

is concerned with the study of how one or more variables affect changes in another variable. It is thus a study of functional relationships existing between two or more variables. This analysis can be termed as regression analysis. Causal analysis is considered relatively more important in experimental researches, whereas in most social and business researches our interest lies in understanding and controlling relationships between variables then with determining causes *per se* and as such we consider correlation analysis as relatively more important.

A small red triangle pointing to the right, serving as a bullet point.

(a) *Multiple regression analysis*: This analysis is adopted when the researcher has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about the dependent variable based on its covariance with all the concerned independent variables.

(b) *Multiple discriminant analysis*: This analysis is appropriate when the researcher has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute. The object of this analysis happens to be to predict an entity's possibility of


belonging to a particular group based on several predictor variables.

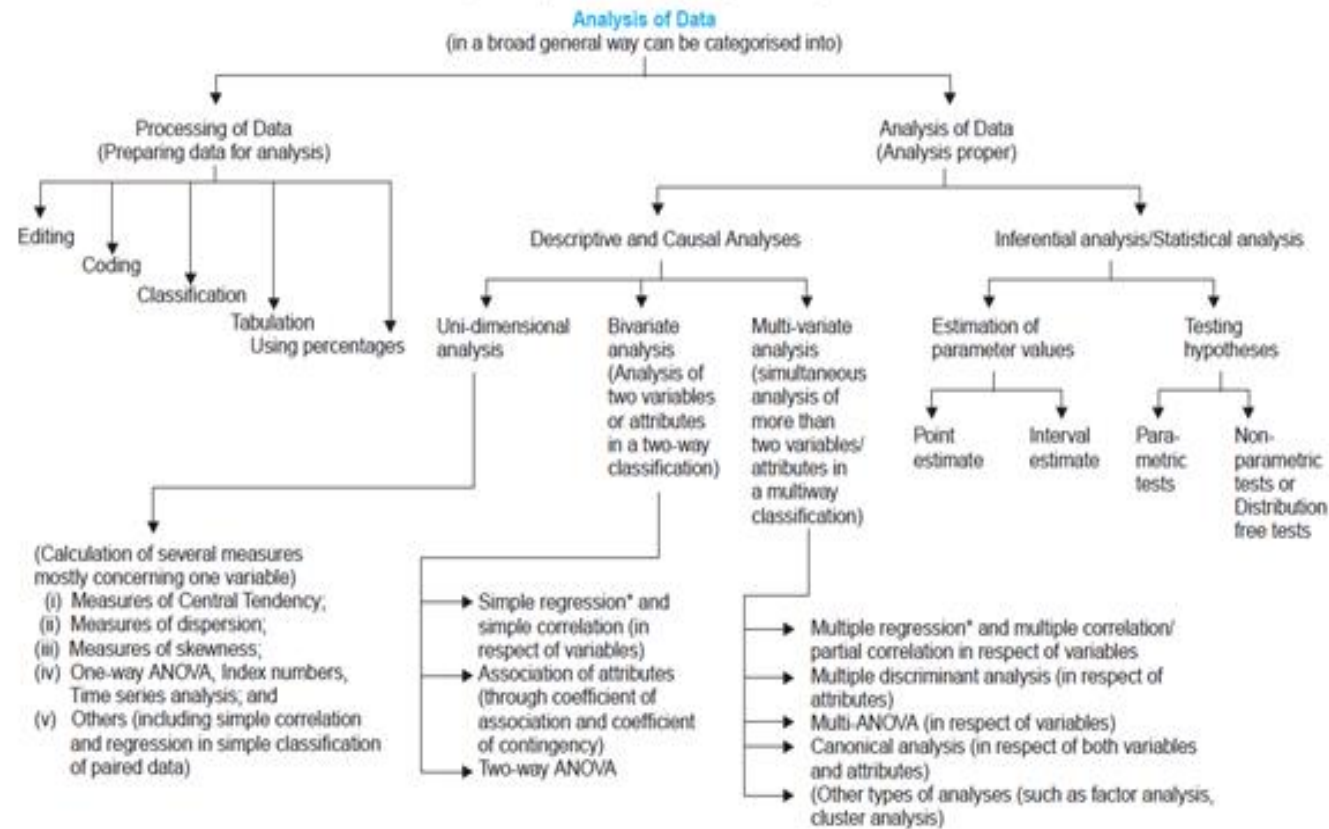
(c) *Multivariate analysis of variance (or multi-ANOVA)*: This analysis is an extension of two way ANOVA, wherein the ratio of among group variance to within group variance is worked out on a set of variables.

(d) *Canonical analysis*: This analysis can be used in case of both measurable and non-measurable variables for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.

Inferential analysis is concerned with the various tests of significance for testing hypotheses in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential

analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

- 
- (1) measures of central tendency or statistical averages;
 - (2) measures of dispersion;
 - (3) measures of asymmetry (skewness);
 - (4) measures of relationship; and
 - (5) other measures.



* Regression analysis (whether simple or multiple) is termed as Causal analysis whereas correlation analysis indicates simply co-variation between two or more variables.

