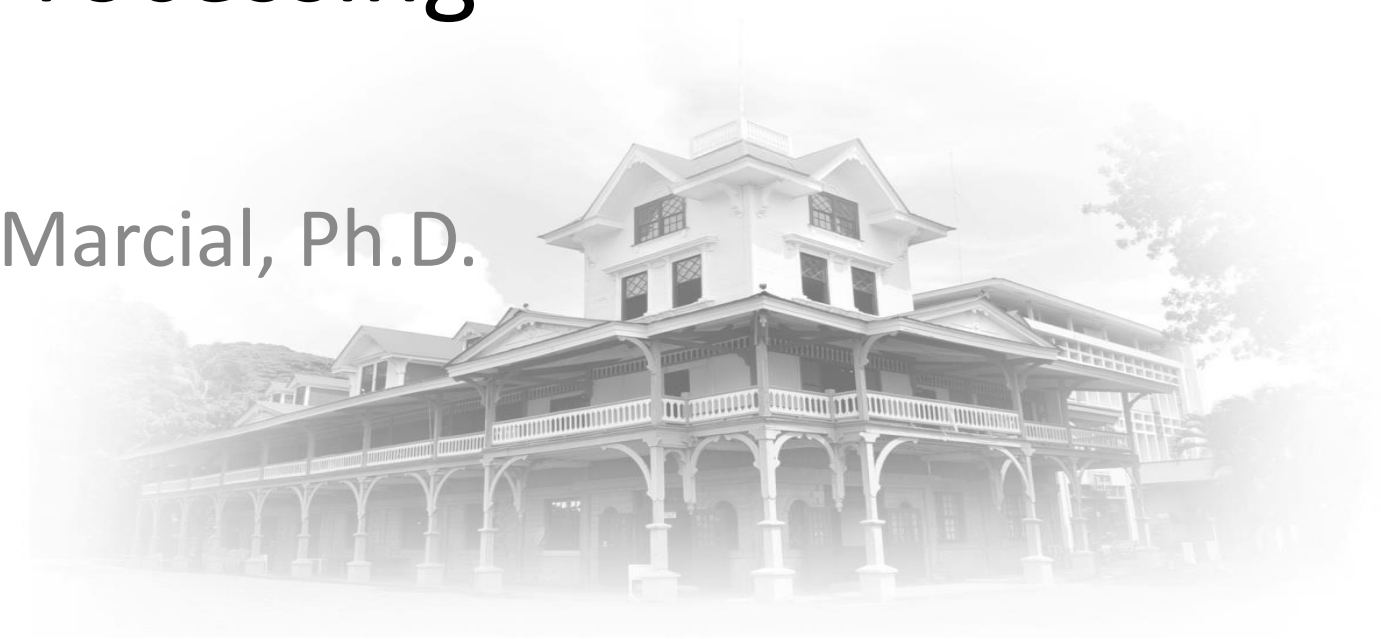
# Data Processing

Dave E. Marcial, Ph.D.



# Data processing

- A series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.

https://sisu.ut.ee/rdm_course1/data-collection-processing-and-analysis

Methods of processing
must be rigorously documented
to ensure
the utility and integrity of the data.

https://sisu.ut.ee/rdm_course1/data-collection-processing-and-analysis

**Data processing** is concerned with editing, coding, classifying, tabulating and charting and diagramming data.

The essence of **data processing** is data reduction. Data reduction involves winnowing out the irrelevant from the relevant data and establishing order from chaos and giving shape to a mass of data.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Classification of Data processing

**1. Business Data processing (BDP)**

- Business data processing is characterized by the need to establish, retain, and process files of data for producing useful information.
- Generally, it involves a large volume of input data, limited arithmetical operations, and a relatively large volume of output.
- For example, a large retail store must maintain a record for each customer who purchases on account, update the balance owned on each account, and a periodically present a bill to the customer for merchandise purchased. This type of record keeping requires reading a customer's account number, name, address, and previous balance. The bill involves a few basic calculations and the result are printed and mailed to the customer for collection. Tens of thousands of similar bills are commonly handled in the same way.

https://www.mbaknol.com/management-information-systems/areas-of-data-processing/

# Classification of Data processing

**2. Scientific Data Processing (SDP)**

- In science, data processing involves a limited volume of input and many logical or arithmetic calculations.
- Unlike business problems, most of the scientific problems are non-repetitive, requiring a “one-time” solution.
- For example, in cancer research, data on cancer patients (collected over a period of time) are analyzed by a computer to produce a possible cure. Although a final cure is unavailable, computer analysis of the hundreds of man-years of computations. It has also brought us a step closer to the final answer to the cancer horror. Although scientific data may differ from business data, the processing pattern is quite similar.

https://www.mbaknol.com/management-information-systems/areas-of-data-processing/

# Data processing consists of five important steps:

## 1. Editing of Data

- Editing is the first step in data processing. Editing is the process of examining the data collected in questionnaires/schedules to detect errors and omissions and to see that they are corrected and the schedules are ready for tabulation. When the whole data collection is over a final and a thorough check up is made.
- Data processor must ensure the following:

1. Accurate as possible,
2. Consistent with other facts secured,
3. Uniformly entered,
4. As complete as possible,
5. Acceptable for tabulation and arranged to facilitate coding tabulation.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 1. Editing of Data

- Types of Editing

**a)** **Editing for quality** asks the following questions: are the data forms complete, are the data free of bias, are the recordings free of errors, are the inconsistencies in responses within limits, are there evidences to show dishonesty of enumerators or interviewers and are there any wanton manipulation of data.

**b)** **Editing for tabulation** does certain accepted modification to data or even rejecting certain pieces of data in order to facilitate tabulation. or instance, extremely high or low value data item may be ignored or bracketed with suitable class interval.

**c)** **Field Editing** is done by the enumerator (in the case of manual data collection). The schedule filled up by the enumerator or the respondent might have some abbreviated writings, illegible writings and the like. These are rectified by the enumerator. This should be done soon after the enumeration or interview before the loss of memory. The field editing should not extend to giving some guess data to fill up omissions.

**d)** **Central Editing** is done by the researcher after getting all schedules or questionnaires or forms from the enumerators or respondents. Obvious errors can be corrected. For missed data or information, the editor may substitute data or information by reviewing information provided by likely placed other respondents. A definite inappropriate answer is removed and “no answer” is entered when reasonable attempts to get the appropriate answer fail to produce results.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 1. Editing of Data

- Types of Editing

**a)** **Editing for quality** asks the following questions: are the data forms complete, are the data free of bias, are the recordings free of errors, are the inconsistencies in responses within limits, are there evidences to show dishonesty of enumerators or interviewers and are there any wanton manipulation of data.

**b)** **Editing for tabulation** does certain accepted modification to data or even rejecting certain pieces of data in order to facilitate tabulation. or instance, extremely high or low value data item may be ignored or bracketed with suitable class interval.

**c)** **Field Editing** is done by the enumerator (in the case of manual data collection). The schedule filled up by the enumerator or the respondent might have some abbreviated writings, illegible writings and the like. These are rectified by the enumerator. This should be done soon after the enumeration or interview before the loss of memory. The field editing should not extend to giving some guess data to fill up omissions.

**d)** **Central Editing** is done by the researcher after getting all schedules or questionnaires or forms from the enumerators or respondents. Obvious errors can be corrected. For missed data or information, the editor may substitute data or information by reviewing information provided by likely placed other respondents. A definite inappropriate answer is removed and “no answer” is entered when reasonable attempts to get the appropriate answer fail to produce results.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 2. Data Coding

- 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 pre-code the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires.
- Coding is the process/operation by which data/responses are organized into classes/categories and numerals or other symbols are given to each item according to the class in which it falls.
- In other words, coding involves two important operations;
  - (a) deciding the categories to be used and
  - (b) allocating individual answers to them.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 2. Data Coding

- The study of the responses is the first step in coding.
- Secondly, coding frame is developed by listing the possible answers to each question and assigning code numbers or symbols to each of them which are the indicators used for coding. The coding frame is an outline of what is coded and how it is to be coded. That is, a coding frame is an outline of what is coded and how it is to be coded. That is, coding frame is a set of explicit rules and conventions that are used to base classification of observations variable into values which are which are transformed into numbers.
- Thirdly, after preparing the sample frame the gradual process of fitting the answers to the questions must be begun.
- Lastly, transcription is undertaken i.e., transferring of the information from the schedules to a separate sheet called transcription sheet. Transcription sheet is a large summary sheet which contain the answer/codes of all the respondents. Transcription may not be necessary when only simple tables are required and the number of respondents are few.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 3. Classification of Data

❖ Objectives of Classification are below:

- The complex scattered and haphazard data is organized into concise, logical and intelligible form.
- It is possible to make the characteristics of similarities and dis – similarities clear.
- Comparative studies is possible.
- Understanding of the significance is made easier and thereby good deal of human energy is saved.
- Underlying unity amongst different items is made clear and expressed.
- Data is so arranged that analysis and generalization becomes possible.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 3. Classification of Data

- Classification is of two types, viz., quantitative classification, which is on the basis of variables or quantity and qualitative classification, in which classification according to attributes.
- The former is the way of, grouping the variables, say, quantifying the variables in cohesive groups, while the latter groups 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 presence or absence of a certain quality.
- Grouping the workers of a factory under various income (class intervals) groups come under the multiple classification; and making two groups into skilled workers and unskilled workers is the dichotomous classification. The tabular form of such classification is known as statistical series, which may be inclusive or exclusive.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 4. Tabulation of data

- Tabulation is the process of summarizing raw data and displaying it in compact form for further analysis.
- Therefore, preparing tables is a very important step. Tabulation may be by hand, mechanical, or electronic.
- The choice is made largely based on the size and type of study, alternative costs, time pressures, and the availability of computers, and computer programmes. If the number of questionnaire is small, and their length short, hand tabulation is quite satisfactory.
- Table may be divided into: (i) Frequency tables, (ii) Response tables, (iii) Contingency tables, (iv) Uni-variate tables, (v) Bi-variate tables, (vi) Statistical table and (vii) Time series tables.
- Generally a data table has the following parts: (a) table number, (b) title of the table, (c) caption (d) stub (row heading), (e) body, (f) head note, (g) foot note.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 4. Tabulation of data

- As a general rule the following steps are necessary in the preparation of table:
  - ✓ Title of table: The table should be first given a brief, simple and clear title which may express the basis of classification.
  - ✓ Columns and rows: Each table should be prepared in just adequate number of columns and rows.
  - ✓ Captions and stubs: The columns and rows should be given simple and clear captions and stubs.
  - ✓ Ruling: Columns and rows should be divided by means of thin or thick rulings.
  - ✓ Arrangement of items; Comparable figures should be arranged side by side.
  - ✓ Deviations: These should be arranged in the column near the original data so that their presence may easily be noted.
  - ✓ Size of columns: This should be according to the requirement.
  - ✓ Arrangements of items: This should be according to the problem.
  - ✓ Special emphasis: This can be done by writing important data in bold or special letters.
  - ✓ Unit of measurement: The unit should be noted below the lines.
  - ✓ Approximation: This should also be noted below the title.
  - ✓ Foot – notes: These may be given below the table.
  - ✓ Total: Totals of each column and grand total should be in one line.
  - ✓ Source : Source of data must be given. For primary data, write primary data.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data processing consists of five important steps:

## 5. Data Diagrams

- Diagrams are charts and graphs used to present data. These facilitate getting the attention of the reader more. These help presenting data more effectively. Creative presentation of data is possible.
- The data diagrams classified into:
  - Charts: A chart is a diagrammatic form of data presentation. Bar charts, rectangles, squares and circles can be used to present data. Bar charts are uni-dimensional, while rectangular, squares and circles are two-dimensional.
  - Graphs: The method of presenting numerical data in visual form is called graph, A graph gives relationship between two variables by means of either a curve or a straight line. Graphs may be divided into two categories. (1) Graphs of Time Series and (2) Graphs of Frequency Distribution. In graphs of time series one of the factors is time and other or others is / are the study factors. Graphs on frequency show the distribution of by income, age, etc. of executives and so on.

https://www.mbaknol.com/research-methodology/methods-of-data-processing-in-research/

# Data Processing Cycle

- The Data Processing Cycle is a series of steps carried out to extract information from raw data. Although each step must be taken in order, the order is cyclic. The output and storage stage can lead to the repeat of the data collection stage, resulting in another cycle of data processing. The cycle provides a view on how the data travels and transforms from collection to interpretation, and ultimately, used in effective business decisions.
- There are 6 stages of data processing cycle:
  1. Collection
  2. Preparation
  3. Input
  4. Processing
  5. Output & Interpretation
  6. Storage

# Data Processing System

- It is a combination of machines and people that produces a defined set of outputs for a certain set of inputs. The inputs and outputs are interpreted as data, facts, information, depending on the interpreter's relation to the system.
- A system may involve some combination of:
  - Conversion is converting data to another format.
  - Validation – Ensuring that supplied data is "clean, correct and useful."
  - Sorting – "arranging items in some sequence and/or in different sets."
  - Summarization – reducing detail data to its main points.
  - Aggregation – combining multiple pieces of data.
  - Analysis – the "collection, organization, analysis, interpretation and presentation of data.".
  - Reporting – list detail or summary data or computed information.
  - Presentation – data presentation is helpful in taking decisions

# Data Processing Methods

**1. Batch Processing.**

- Batch processing is a technique in which data to be processed or programs to be executed are collected into groups to permit convenient, efficient, and serial processing.
- It is the simplest form of data processing.
- With this method, data is entered to the information flow in large volumes, or batches. That is, the processing by computer is performed periodically, at specified time intervals (weekly, monthly, etc) when large volumes are accumulated.
- Daily transactions in a business establishment, for example, may be batch processed on a weekly basis. Instead of being processed periodically when a sufficient volume has been accumulated.
- Advantages of batch processing are:
  - Economical when a large volume of data must be processed and
  - The most appropriate method for those applications (e.g., payroll) where the delay caused by accumulating data into batches does not reduce the value of the information.
- Limitations of batch processing are:
  - It requires sorting prior to processing
  - Reduce timeliness in some instances and
  - Requires sequential file organization.

https://www.mbaknol.com/management-information-systems/data-processing-methods/

# Data Processing Methods

**2. On-line Processing.**

- The term “online” refers to equipment or devices under the direct control of the central processing unit (CPU) of a computer.
- An on-line operation, then, is one which uses devices directly connected to the CPU either for data entry or inquiry purposes.
- That is, with a terminal we can either enter data or inquire about the status of some record or file that is stored by the computer.

https://www.mbaknol.com/management-information-systems/data-processing-methods/

# Data Processing Methods

## 3. Real-time Processing

- Real- time processing is a method of data processing which has the capability of a fast-response to obtain data from an activity or a physical process, perform computations, and return a response rapidly enough to affect the outcome of the activity or process. In order words, the processing of data is done immediately.
- Airline reservation systems, for example, require immediate processing. Each time a ticket is issued or cancelled, or a plane's schedule is altered, the data must be immediately entered sing is also used in keeping track of the availability of motel and hotel rooms, and in immediate updating of customer records in saving banks.

https://www.mbaknol.com/management-information-systems/data-processing-methods/

# Data Processing Methods

## 4. Distributed Processing

- The most complex level of computer processing, distributed processing, generally consists of remote terminals linked to a large central computer system to help the user conduct inquiries about accounts, process jobs, or other data processing operations.
- Distributed computer-communications network is similar in some respects to public utilities such as telephone and electric companies — e.g., electric power plants are geographically dispersed and the energy resources generated are transmitted through a coordinating regional network or grid to the places where the energy resources are needed.
- Some of the advantages of distributed processing system are:
  - central processor idle time is reduced.
  - sophisticated computers and a growing library of applications programs may be immediately available to end-users whenever needed.
  - skilled professionals are available to help users develop their own specialized applications.
  - managers may be able to react more rapidly to new developments and interact with the system in order to seek solutions to unusual problems.

https://www.mbaknol.com/management-information-systems/data-processing-methods/

# Data Processing Methods

## 4. Distributed Processing

- The possible disadvantages are:
  - the reliability and the cost of data communication facilities used, and the cost and quality of the computing service received, may be disappointing in some cases.
  - input/output terminals are often rather slow and inefficient.
  - provisions for protecting the confidentially and integrity of user programs and data files are generally ineffective against a skilled penetrator.

https://www.mbaknol.com/management-information-systems/data-processing-methods/

# Data Processing Methods

## 5. Multi Processing

- This type of processing perhaps the most widely used types of data processing.
- It is used almost everywhere and forms the basic of all computing devices relying on processors.
- Multi processing makes use of CPUs (more than one CPU).
- The task or sets of operations are divided between CPUs available simultaneously thus increasing efficiency and throughput.
- The break down of jobs which needs be performed are sent to different CPUs working parallel within the mainframe.
- The result and benefit of this type of processing is the reduction in time required and increasing the output .
- Moreover CPUs work independently as they are not dependent on other CPU, failure of one CPU does not result in halting the complete process as the other CPUs continue to work.
- Examples include processing of data and instructions in computer, laptops, mobile phones etc.

# Data Processing Methods

## 6. Time sharing

- Time based used of CPU is the core of this data processing type.
- The single CPU is used by multiple users.
- All users share same CPU but the time allocated to all users might differ.
- The processing takes place at different intervals for different users as per allocated time.
- Since multiple users can uses this type it is also referred as multi access system.
- This is done by providing a terminal for their link to main CPU and the time available is calculated by dividing the CPU time between all the available users as scheduled.