INTRODUCTION

Collection of data and display of results is very important in any study. The data of an experimental study, observational study or a survey are required to be collected in properly designed format for documentation, taking into consideration the design of study and different end points of the study. Usually data are collected in the proforma of the study. The data recorded and documented should be stored carefully in documents and in electronic form for example, excel sheets or data bases.

The data are usually classified into qualitative and quantitative [Table 1]. Qualitative data is further divided into two categories, unordered qualitative data, such as blood groups (A, B, O, AB); and ordered qualitative data, such as severity of pain (mild, moderate, severe). Quantitative data are numerical and fall into two categories: discrete quantitative data, such as the internal diameter of endotracheal tube; and continuous quantitative data, such as blood pressure.[1]

Data Coding is needed to allow the data recorded in categories to be used easily in statistical analysis with a computer. Coding assigns a unique number to each possible response. A few statistical packages analyse categorical data directly. If a number is assigned to categorical data, it becomes easier to analyse. This means that when the data are analysed and reported, the appropriate label needs to be assigned back to the numerical value to make it meaningful. The codes such as 1/0 for yes/no has the added advantage that the variable’s 1/0 values can be easily analysed. The record of the codes modified is to be stored for later reference. Such coding can also be done for categorical ordinal data to convert in to numerical ordinal data, for example the severity of pain mild, moderate and severe into 1, 2 and 3 respectively.

PROCESS OF DATA CHECKING, CLEANING AND EDITING

In clinical research, errors occur despite designing the study properly, entering data carefully and preventing errors. Data cleaning and editing are carried out to identify and correct these errors, so that the study results will be accurate.[2]
Data entry errors in case of sex, dates, double entries and unexpected results are to be corrected unquestionably. Data editing can be done in three phases namely screening, diagnosing and editing [Figure 1].

**Screening phase**
During screening of data, it is possible to distinguish the odd data, excess of data, double entries, outliers, and unexpected results. Screening methods are checking of questionnaires, data validation, browsing the excel sheets, data tables and graphical methods to observe data distribution.

**Diagnostic phase**
The nature of the data can be assessed in this phase. The data entries can be true normal, true errors, outliers, unexpected results.

**Treatment phase**
Once the data nature is identified the editing can be done by correcting, deleting or leaving the data sets unchanged.

The abnormal data points usually have to be corrected or to be deleted. However some authors advocate these data points to be included in analysis. If these extreme data points are deleted, they should be reported as “excluded from analysis”.

### ROLE OF COMPUTERS IN RESEARCH

The role of computers in scientific research is very high; the computers have the ability to perform the analytic tasks with high speed, accuracy and consistency. The Computers role in research process can be explained in different phases.

**Role of computer in conceptual phase**
The conceptual phase consists of formulation of research problem, literature survey, theoretical frame work and developing the hypothesis. Computers are useful in searching the literatures. The references can be stored in the electronic database.

**Role of computers in design and planning phase**
This phase consists of research design preparation and determining sample design, population size, research variables, sampling plan, reviewing research plan and pilot study. The role of computers in these process is almost indispensable.

**Role of computers in data collection phase**
The data obtained from the subjects stored in computers are word files or excel spread sheets or statistical software data files or from data centers of hospital information management systems (data warehouse). If the data are stored in electronic format checking the data becomes easier. Thus, computers help in data entry, data editing, and data management including follow up actions. Examples of editors are Word Pad, SPSS data editor, word processors.

**Role of computers in data analysis**
This phase mainly consist of statistical analysis of the data and interpretation of results. Software like Minitab (Minitab Inc. USA.), SPSS (IBM Crop. New York), NCSS (LLC. Kaysville, Utah, USA) and spreadsheets are widely used.

**Role of computer in research publication**
Research article, research paper, research thesis or research dissertation is typed in word processing software in computers and stored. Which can be easily published in different electronic formats.

### DATA DISPLAY AND DESCRIPTION OF RESEARCH DATA

Data display and description is an important part of any research project which helps in knowing the distribution of data, detecting errors, missing values and outliers. Ultimately the data should be more comprehensible and meaningful.

Tables are commonly used for describing both qualitative and quantitative data. The graphs are...
useful for visualising the data and understanding the variations and trends of the data. Qualitative data are usually described by using bar or pie charts. Histograms, polygons or box plots are used to represent quantitative data.\(^{[1]}\)

**Qualitative data**

**Tabulation of qualitative data**

The qualitative observations are categorised in to different categories. The category frequency is nothing but the number of observations with in that category. The category relative frequency can be calculated by dividing the number of observations in the category by total number of observations. The Percentage for a category is more commonly used to describe qualitative data. It can be computed by multiplying relative frequency with hundred.\(^{[6,7]}\)

The classification of 30 Patients of a group by severity of postoperative pain presented in Table 2. The frequency table for this data computed by using the software NCSS\(^{[8]}\) is shown in Table 3.

**Graphical display of qualitative data**

The qualitative data are commonly displayed by bar graphs and pie charts.\(^{[9]}\)

Bar graphs displays information of the frequency, relative frequency or percentage of each category on vertical axis or horizontal axis of the graph. \([Figure 2]\) Pie charts depicts the same information in divided slices in a complete circle. The area for the circle is equal to the frequency, relative frequency or percentage of that category \([Figure 3]\).

**Quantitative data**

**Tabulation of quantitative data**

The quantitative data are usually presented as frequency distribution or relative frequency rather than percentage. The data are divided into different classes. The upper and lower limits or the width of classes will depend up on the size of the data and can easily be adjusted.

The frequency distribution and relative frequency distribution table can be constructed in the following manner:

1. The quantitative data are divided into number of classes. The lower limit and upper limit of the classes have to be defined.

| Post-operative Pain | Severity of Pain |
|---------------------|------------------|
| 1                   | Moderate         |
| 2                   | Moderate         |
| 3                   | Mild             |
| 4                   | Moderate         |
| 5                   | Severe           |
| 6                   | Severe           |
| 7                   | Mild             |
| 8                   | Moderate         |
| 9                   | Mild             |
| 10                  | Moderate         |
| 11                  | Mild             |
| 12                  | Moderate         |
| 13                  | Severe           |
| 14                  | Mild             |
| 15                  | Moderate         |
| 16                  | Moderate         |
| 17                  | Moderate         |
| 18                  | Mild             |
| 19                  | Moderate         |
| 20                  | Severe           |
| 21                  | Severe           |
| 22                  | Mild             |
| 23                  | Moderate         |
| 24                  | Mild             |
| 25                  | Moderate         |
| 26                  | Mild             |
| 27                  | Moderate         |
| 28                  | Severe           |
| 29                  | Mild             |
| 30                  | Moderate         |

**Table 3: The frequency table for the variable pain**

| Severity of Pain | Frequency | Cumulative frequency | Percentage | Cumulative percentage |
|------------------|-----------|----------------------|------------|-----------------------|
| Mild             | 10        | 10                   | 33.33      | 33.33                 |
| Moderate         | 14        | 24                   | 46.66      | 80                    |
| Severe           | 6         | 30                   | 20         | 100                   |

\(\text{Figure 2: The bar graph generated by computer using NCSS software for the variable pain}\)
2. The range or width of the class intervals can be calculated by dividing the difference in the upper limit and lower limit by total number of classes.
3. The class frequency is the number of observations that fall in that class.
4. The relative class frequency can be calculated by dividing class frequency by total number of observations.

Example of frequency table for the data of Systolic blood pressure of 60 patients undergoing craniotomy is shown in Table 4. The number of classes were 20, the lower limit and the upper limit were 86 mm of Hg and 186 mm of Hg respectively.

Graphical description of quantitative data

**Histogram**
The frequency distribution is usually depicted in histograms. The count or frequency is plotted along the vertical axis and the horizontal axis represents data values. The normality of distribution can be assessed visually by histograms. A frequency histogram is constructed for the dataset of systolic blood pressure, from the frequency Table 4 [Figure 4].

**Box plots**
Box plot gives the information of spread of observations in a single group around a centre value.

| Class      | Lower limit | Upper limit | Midpoint | Frequency | Relative frequency | Cumulative frequency | Cumulative relative frequency |
|------------|-------------|-------------|----------|-----------|--------------------|-----------------------|-------------------------------|
| At or below| 86.0        | 0           | 0        | 0.00      | 0.00               | 0                     | 0.0000                        |
| 1          | 86.0        | 91.56       | 88.78    | 1         | 0.05               | 1                     | 0.017                         |
| 2          | 91.56       | 97.11       | 94.33    | 3         | 0.05               | 4                     | 0.067                         |
| 3          | 97.11       | 102.67      | 99.89    | 1         | 0.05               | 5                     | 0.08                          |
| 4          | 102.67      | 108.22      | 105.44   | 3         | 0.05               | 8                     | 0.13                          |
| 5          | 108.22      | 113.78      | 111.0    | 1         | 0.05               | 9                     | 0.15                          |
| 6          | 113.78      | 119.33      | 116.56   | 8         | 0.13               | 17                    | 0.28                          |
| 7          | 119.33      | 124.89      | 122.11   | 3         | 0.05               | 20                    | 0.33                          |
| 8          | 124.89      | 130.44      | 127.67   | 15        | 0.25               | 35                    | 0.58                          |
| 9          | 130.44      | 136.0       | 132.22   | 4         | 0.07               | 39                    | 0.65                          |
| 10         | 136.0       | 141.56      | 138.78   | 5         | 0.08               | 44                    | 0.73                          |
| 11         | 141.56      | 147.11      | 144.33   | 6         | 0.10               | 50                    | 0.83                          |
| 12         | 147.11      | 152.67      | 149.89   | 3         | 0.05               | 53                    | 0.88                          |
| 13         | 152.67      | 158.22      | 155.44   | 6         | 0.10               | 59                    | 0.98                          |
| 14         | 158.22      | 163.78      | 161.0    | 0         | 0.00               | 59                    | 0.98                          |
| 15         | 163.78      | 169.33      | 166.56   | 0         | 0.00               | 59                    | 0.98                          |
| 16         | 169.33      | 174.89      | 172.11   | 1         | 0.05               | 60                    | 1.00                          |
| 17         | 174.89      | 180.44      | 177.67   | 0         | 0.00               | 60                    | 1.00                          |
| 18         | 180.44      | 186.0       | 183.22   | 0         | 0.00               | 60                    | 1.00                          |
| Above      | 186.0       | 0           | 0        | 0.00      | 60                 | 1.00                  |                                |

**Figure 3:** The Pie graph generated by computer using NCSS software for the variable pain

**Figure 4:** The frequency histogram for the data set of systolic blood pressure (BP), for which the frequency table is constructed in Table 4
The distribution pattern and extreme values can be easily viewed by box plot. A boxplot is constructed for the dataset of systolic blood pressure, from the frequency Table 4 [Figure 5].

**Polygons**

Polygon construction is similar to histogram. However it is a line graph connecting the data points at mid points of class intervals. The polygon is simpler and outline the data pattern clearly [Figure 6].

It is often necessary to further summarise quantitative data, for example, for hypothesis testing. The most important elements of a data are its location, which is measured by mean, median and mode. The other parameters are variability (range, interquartile range, standard deviation and variance) and shape of the distribution (normal, skewness, and kurtosis). The details of which will be discussed in the next chapter.

**SUMMARY**

The proper designing of research methodology is an important step from the conceptual phase to the conclusion phase and the computers play an invaluable role from the beginning to the end of a study. The data collection, data storage and data management are vital for any study. The data display and interpretation will help in understating the behaviour of the data and also to know the assumptions for statistical analysis.

**REFERENCES**

1. Whitley E, Ball J. Statistics review 1: Presenting and summarising data. Crit Care 2002;6:66-71.
2. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: Detecting, diagnosing, and editing data abnormalities. PLoS Med 2005;2:e267.
3. Gardner MJ, Altman DG. Statistics with confidence. London: BMJ; 1994. p. 140.
4. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post randomisation exclusions: The intention to treat principle and excluding patients from analysis. BMJ 2002;325:652-4.
5. Chamanlal J. Role of computer applications and tools in the scientific research process. Int J Res Sci Technol 2014;3:33-44.
6. Hall RO. Handbook of Tabular Presentation. New York: The Ronald Press Company; 1943.
7. Nicol AA, Pexman MP. Presenting Your Findings: A Practical Guide for Creating Tables. Washington, DC: American Psychological Association; 1999.
8. NCSS 10 Statistical Software 2015. NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/ncss.
9. Wainer H. Understanding graphs and tables. Educ Res 1992;21:14-23.