

## BIostatISTICS INTRODUCTION AND DATA REPRESENTATION USED IN DENTISTRY DEPARTMENT

Amir Bukhari \*  
Mehboob Ali Shah \*\*

### ABSTRACT

*During recent decades increased emphasis has been laid down on the development and use of statistical techniques for the analysis of problems. With the result these methods are widely used in research.<sup>(1)</sup>*

*Some significant details of the Biostatistics have been mentioned. The result has to do with the details of construction of diagrams and charts. This research has practical significance and some statements about the patients are given. Selection of an appropriate sample size is one of the most important aspects of any statistical problem.<sup>(1)</sup>*

### INTRODUCTION

To day the modern sciences need a free and adequate flow of information. To make better decisions this information should be analyzed using methods and techniques, specially developed for this purpose and are covered in the field known as "Statistics".<sup>(9)</sup>

### STATISTICS

In general people talk about statistics as data collection, tabulation and presentation. That is why "Oxford" American Dictionary <sup>(5)</sup> defines the word "Statistics as the services of collection, classifying and interpreting information based on a number of things. But in fact statistics does much more than the way it is defined.<sup>(4)</sup> In simple words the statisticians not only collect, classify and tabulate the data, but they also analyze it to make some generalization and decision about it. For example the scientists, the industrialists and the agriculturists, the

---

\* Assistant Professor, Faculty of Dentistry, Liaquat University of Medical and Health Sciences, Jamshoro

\*\* Professor, Department of Statistics, University of Sindh, Jamshoro

administrators, the educationists the social workers all are substantially benefited by the use of statistics.

Statistics are effective in the studies of population changes, birth and death rates and incidence of disease.

### COLLECTION OF DATA

The collection of figures is the very first step in an inquiry. The data to be collected are generally of two types:

1. Primary data
2. Secondary data.

#### 1. PRIMARY DATA

Primary data are those original data which are collected and recorded by the investigator and are utilized as raw material for his inquiry or for his own use.

#### 2. SECONDARY DATA

Secondary data are those which are collected by another person but are utilized by the investigator for his own use. These are collected, classified, tabulated and presented in a suitable form to serve the required purposes.

### WHAT IS TABULATION?

Tabulation is the presentation of classified data in a systematic and scientific way and in a form so that the special significance of the data is clear. It is the process of presenting and arranging the data in an orderly manner in tables.<sup>(3)</sup>

### BIO-STATISTICS

When the data being analyzed are taken from one of the health sciences we use the term bio-statistics or bio-metry (bio stands for biology and metry stands for measurements).<sup>(7)</sup>

Because numerical data are the raw material of statistical investigation, one of the first steps in any statistical study must be the collection of suitable data.<sup>(2)</sup> The only step that ordinarily precedes the collection of data is the careful planning of the study and the precise formulation of its purpose, scope and objectives.

When dealing with large sets of data we are often more interested in knowing the percentage of the total number of cases that fall into a given class, than in knowing the actual class frequency. If this is the case we can easily convert a given distribution into a percentage distribution by dividing each class frequency by the total number of cases and then multiplying by 100, if omitted the last step and did not multiply by 100, we would call the resulting distribution of proportions instead of a percentage distribution.

#### DIAGRAMS AND GRAPHS

A diagram or graph is a method of presenting data in visual form. It is the pictorial representation of data in visual form. It is the pictorial representation of data for studying changes in a single variable or for making comparisons between several similar or related variables.

Diagrams and graphs are very useful. The advantage of presentation of data by charts and graphs is that the significant the salient features stand out prominently and immediately.<sup>(2)</sup>

#### GROUPED DATA THE FREQUENCY DISTRIBUTION

A set of observations can be made more comprehensive and meaning full by means of an ordered array further useful summarization may be achieved by grouping the data. To group a set of observations we select a set of continuous, non-overlapping intervals such that each value in the set of observations can be placed in one and only one of the intervals. These intervals are usually referred to as class interval.<sup>(8)</sup>

**Examples**

Table I shows the weights in ounces of malignant tumors removed from the abdomens of 57 subjects.

**TABLE - I**  
**WEIGHTS IN OUNCES OF MALIGNANT TUMORS REMOVED**  
**FROM THE ABDOMENS OF 57 SUBJECTS**

68	24	28	27	22	36
63	25	25	31	43	28
42	44	45	50	27	31
27	65	12	38	49	22
30	43	57	21	28	23
36	25	51	16	23	30
28	74	12	24	19	12
32	51	32	69	46	
79	42	49	47	43	
27	42	38	23	43	

We may now construct our intervals. Since the smallest value in the table is 12 and the largest value 79 we may begin our intervals with 10 and with 70.

Determining the number of values falling into each class interval is merely a matter of looking at the values one by one and of placing a tally mark beside the appropriate interval when we do this we have Table - II.

**TABLE - II**  
**FREQUENCY DISTRIBUTION OF WEIGHTS (IN OUNCES)**

Frequency Distribution of weights (in ounces)		
Class interval	Tally	Frequency
10 - 19		5
20 - 29		19
30 - 39		10
40 - 49		13
50 - 59		4
60 - 69		4
70 - 79		2
<b>Total</b>		<b>57</b>

A table of this kind is called a frequency distribution. This table shows the way in which the values of the variable are distributed among the specified class intervals. By consulting it, we can determine the frequency of occurrence of values within any one of the class intervals.

### THE HISTOGRAM

We may display a frequency distribution graphically in the form of a histogram as shown in figure 1. In constructing a histogram the values of the variable under consideration make up the horizontal axis, while the vertical axis has as its scale the frequency of occurrence. Above each class interval on the horizontal axis a rectangular bar or call as it is some times called is erected so that the height corresponds to the respective frequency. The calls of a histogram must be joined and to accomplish.

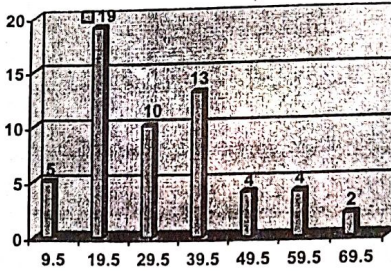
This we must take into account the true boundaries of the class intervals to prevent gaps from occurring between the cells of our graph.

TABLE - III

THE DATA OF TABLE - 2 SHOWING TRUE CLASS LIMITS

The data of table 2 showing true class limits		
Class Limits	Tally	Frequency
9.5 - 19.5		5
19.5 - 29.5		19
29.5 - 39.5		10
39.5 - 49.5		13
49.5 - 59.5		4
59.5 - 69.5		4
69.5 - 79.5		2

FIGURE - I

**THE FREQUENCY POLYGON**

A frequency distribution can be portrayed graphically in yet another way by means of a frequency polygon. To draw a frequency polygon we first place a dot above the mid point of each class interval represented on the horizontal axis of a graph like the one shown in figure 1. The height above the horizontal axis of a given dot corresponds to the frequency of the relevant class interval. Connecting the dots by straight lines produces the frequency polygon.

FIGURE 2 IS THE FREQUENCY POLYGON FOR THE DATA OF EXAMPLE - 1

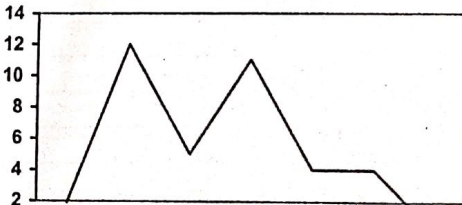


Fig - 2 Frequency polygon of weights

### THE ORDERED ARRAY

When measurements of a random variable are taken on the entities of a population, the resulting values usually come to the researcher as a mass of unordered data. Unless the number of observations is extremely small it will run likely that these data will impart much information until they have been put in some kind of order. If the number of observations is not too large, a first step in organizing these data is the preparation of an ordered array. An ordered array is a listing of the values of a collection in order of magnitude from the smallest value to the largest values.

#### Example

Table – IV consists of a list of the ages of patients admitted to a Dental Hospital during a certain day. The table reflects the order in which the patients were admitted. As can be seen it requires considerable searching to ascertain such elementary information as the age of the youngest and oldest patient admitted. Table – V present the same date in the form of an ordered array.

Raw data from Dentistry Department:

TABLE – IV

AGE (IN YEARS) OF PATIENTS ADMITTED TO A DENTAL HOSPITAL DURING A CERTAIN DAY

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Age	10	22	24	42	37	77	89	85	28	63	9	10	7	51	2

TABLE – V

AN ORDERED ARRAY OF THE AGES OF PATIENTS ADMITTED TO A DENTAL HOSPITAL DURING A CERTAIN DAY

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Age	2	7	9	10	10	22	24	28	37	42	51	63	77	85	89

The mean is obtained by adding all the values in a population or sample and dividing by the number of values that are added. To obtain the mean age of the population of 15 patients represented in Table - IV, we proceed as follows:

$$\text{Mean age} = 10 + 22 + \dots + 51 + 2/15 = 37$$

### THE RANGE

One way to measure the variation in a set of a value is to compute the range. The range is the difference between the smallest and largest value in a set of observations. If we denote the range by  $R$ , the largest value by  $X_L$  and the smallest value by  $X_s$ , we compute the range as follows:

$$R = X_L - X_s$$

Using the data in Table - V we have

$$R = 87$$

### THE VARIANCE

When the values of a set of observations lie close to their mean, the dispersion is less than when they are scattered over a wide range. Since this is true, it would be intuitively appealing if we could measure dispersion relative to the scatter of the values about their mean. Such a measure is realized in what is known as the variance, we subtract the mean from each of the values, square the differences and then add them up. This sum of the squared deviations of the values from their mean is divided by the simple size.

### THE COEFFICIENT OF VARIATION

The standard deviation is useful as a measure of variation within a given set of data. When one desired to compare the dispersion in two sets of data, however, fallacious results. What is needed is situations like there is a measure of relative variation rather than absolute variation. Such a measure is found in the coefficient of variation which expresses the standard deviation as a percentage of the mean.

The formula is given by:  $C.V = \frac{S}{X} \times (100)$

**SUPPOSE TWO SAMPLES OF HUMAN MALES YIELD THE FOLLOWING RESULTS**

	<b>Sample – I</b>	<b>Sample – II</b>
Age	25 years	11 years
Mean weight	145 pounds	80 pounds
Standard deviation	10 pounds	10 pounds

A comparison of the standard deviation might lead one to conclude that the two samples possess equal variability. If we compute the coefficients of variation, however, we have for the twenty – five years – olds.

$$C.V = 10/145 (100) = 6.9$$

And for the eleven year – olds

$$C. V = 10/80 (100) = 12.5$$

If we compare these results we get quite a different impression. The coefficient of variation is also useful in comparing the results obtaining by different persons who are conducting investigations involving the same variables.

**SUGGESTION OF COMPUTERS AND BIO-STATISTICAL ANALYSIS**

The relatively recent wide spread use of computers has had a tremendous impact on health in particular. Computers can perform more calculation faster and for more accurately than human technicians. The use of computers makes it possible for investigators to devote more time to the improvement of the quality of raw data and the interpretation of the results.

**REFERENCES**

1. Bernad, G. Greenberg: *Bio-Statistics*, McGraw Hill N. York, 1965.
2. Pearca, S. C.: *Biological Statistics: An Introduction*, McGraw Hill, N. York, 1965.
3. Mather: *Statistical Analysis in Biology*, Chapman Hall 1964.

4. Robert Steven Ledley: *Use of Computers in Biology and Medicine*, McGraw Hill, N. York, 1965.
  5. Wiltred J. Dixon and Frank J. Massey Jr.: *Introduction to Statistical Analysis*, (3<sup>rd</sup> edition) Mc-Graw Hill, New York, 1969.
  6. Harold M. Schoolman: *Statistics in Medical Research*, The New England Journal of Medicine, 280, (1969), 218-A.
  7. Alexander M. Mood, Franklin A. Graybill: *Introduction to the Theory of Statistics*, (3<sup>rd</sup> edition), McGraw Hill, New York, 1974.
  8. Robert R. Sokal and F. Jams Rohlf: *Bimetry* W. H. Freeman, San Francisco, 1969.
  9. *Oxford American Dictionary*, Oxford University Press, New York, 1980.
-