Simplified

Biostatistics

For Nurses

Special Features

- First book on Biostatistics specifically designed for Nurses
- **100+** illustrations including flowcharts, tables, and graphs
- **100+** Solved Examples
- **10+** Appendices containing information on the applicability of biostatistics in research
- Practical boxes with applicable statistical tips
- Exclusively included statistical tools and their applications in the biostatistics

CBS Publishers & Distributors Pvt. Ltd.

Anju Dhir

5

Frequency Distribution

"Clutter is nothing more than postponed decisions."
*—**Barbara Hemphill***

LEARNING OBJECTIVES

After the completion of the chapter, the readers will be able to:

- Understand frequency distribution.
- Make frequency distribution table.
- Make cumulative frequency distribution table.
- Make relative frequency percentage table.

CHAPTER OUTLINE

- Introduction
- Practical Importance of Frequency Distribution
- Frequency Distribution Table

INTRODUCTION

Frequency distribution is a tabular/graphical representation of a data. The counts or frequencies of various outcomes in a sample are presented within a particular interval. Therefore, frequency distribution is a summary of values obtained from a sample data. Frequency distribution shows a compiled view of the entire data in an organized way.

- **Univariate:** The frequency distribution can be univariate when single variable is depicted.
- **Multivariate**: It can be multivariate when more than one variable is depicted.

Let us understand what frequency is?

Frequency: Suppose, we have data of marks obtained by 10 students from a class in a test as given here: 23, 26, 11, 18, 9, 21, 23, 11, 22, 11

This form of data is known as raw data. When we look at the data, it is clear that some digits are repeating. For example, 23 is repeating for 2 times, 11 is repeating for 3 times, whereas other digits are occurring only once. It means the frequency of 23 is 2, and of 11 is 3, whereas for other digits it is 1.

Must Know

- Frequency distribution is used in both qualitative as well as quantitative data.
- Histograms, bar charts, pie charts and line charts are used as frequency distribution charts.

Therefore, we can say that the frequency of a particular data value is the number of times the data value occurs in that particular set of data. The frequency of a data value is represented by 'f'.

PRACTICAL IMPORTANCE OF FREQUENCY DISTRIBUTION

- A well-constructed frequency distribution table makes it possible to analyze the structure of population with respect to a given characteristic.
- The groups into which a population can be broken down, are easily determined by looking at the frequency distribution.
- The nature of distribution of the members of a population with respect to a given characteristic can be ascertained—whether the distribution is symmetric or asymmetric; or the degree of concentration of a particular value.
- Calculation of statistics becomes comfortable like calculation of range of a character or average value, etc.
- We can calculate the degree of skewness, and the measure of kurtosis (the degree of closeness of a cluster of values of characteristics around an average value).
- For easy understanding, a frequency distribution can be represented graphically such as histogram, polygon, etc.

FREQUENCY DISTRIBUTION TABLE

A large data is very difficult to understand and interpret unless it is 'organized'. Here, we have used a term interpret which means 'statistical interpretation' to get the results and to forecast some strategy. To do so, we have to organize data into a concise form so that interpretation and analysis becomes easy. A frequency distribution table talks about the grouping of data of a population with respect to quantitative characteristics. Sometimes, we arrange the values in increasing or decreasing values of magnitude and then it is called 'ranked'. The frequency table is based on some continuously varying characteristic like age, or height, or weight and so on.

- **Ungrouped data:** Here, we calculate the frequency for each observation one by one.

  Example: The test scores of 20 students are as follows (Table 5.1):
  23, 26, 11, 18, 9, 21, 23, 30, 22, 11, 21, 20, 11, 13, 23, 11, 29, 25, 26, 26

  Solution: Some values are appearing more than once in this data that is frequency of few values is more. The table here can help in understanding frequency as:
  In the above example, the frequency of 21 is 2, frequency of 26 and 23 is 3 and the frequency of 11 is 4.

TABLE 5.1: Frequency table for the students marks obtained in the test

Marks obtained in the test	9	11	13	18	20	21	22	23	25	26	29	30	Total
No. of students (frequency)	1	4	1	1	1	2	1	3	1	3	1	1	20

- **Grouped data:** Now consider a situation where we have to collect data for the test scores of 100 students. We will have 100 observations and now it will become difficult to tally for each and every score of all 100 observations. Moreover, the table obtained will be very large in length, will occupy more space and will not be easy to understand. Here, we use a grouped frequency

Practical Tip

The total frequency must always be total of the number of observations after tallying. In the example given above, the total is 20 which is the total number of observations too.

distribution table. Most of the data that we come across in real life is in the form of grouped data. Generally, the amount of data is large and associated with corresponding frequencies of each value and we may divide data items into class intervals to further condense our data. For example, we have a data about hypertension patients of varying age groups in a hospital. The data here can be displayed in classes associated with their corresponding frequencies depending on the number of patients falling in each class interval. Grouped data can be further classified into two types:

1. **Discrete frequency distribution:** Frequency distributions can be grouped like for a discrete characteristic if the range of this particular characteristic is fairly large. Here, the individual data entry is accompanied by its corresponding frequency. There are two columns and in one column we write the individual data items, denoted by **X,** and in other column we write frequencies, denoted by **f.** For example, the distribution of primary health centers in a state with respect to number of patients visiting it.
2. **Continuous frequency distribution:** Here, the data entries are grouped into various class intervals and their corresponding frequencies. There is one column for class intervals and another column for frequencies.

Process of Making Frequency Distribution Table

To make a frequency distribution, several steps are followed:

- **First step is to decide the number of classes:** Neither too less classes nor too many classes are good, as in either case the data will not be presented justifiably. The maximum number of classes can be determined by formula:

  Number of classes (C) = 1 + 3 – 3 log (n)

  Or C = $\sqrt{n}$ (approximately); where n is total number of observations in a data.

- **Second step is to calculate the range of data:** By looking carefully on the values collected in a data.

  Range: It is defined as the difference between the maximum (L) and minimum value (S) of a data. Therefore, range can be calculated as:

  Range = L–S

- **Third step is to calculate the width (h) of a class**:

  h = range/number of classes

  Third step will give class interval or width to condense our huge data.

Practical Tips

- Keep the width of the class interval uniform in all the classes.
- The first class must cover the lowest value of data and the last class must contain the highest value of the data.
- The starting point of first class is arbitrary and may be less than or equal to the minimum value obtained from the data. The midpoint or the average of lower- and upper-class limits of the first-class limits must be perfectly included.
- Keep running tally till the last observation, to keep a check that no observation is missed.

Let us understand with examples:

Example: A survey was taken of 20 families, to find out how many kids they have and the results were recorded as follows:

1, 2, 1, 0, 3, 4, 0, 1, 1, 1, 2, 2, 3, 2, 3, 2, 1, 4, 0, 0

Solution:

- Divide the results (x) into intervals, and then count the number of results in each interval. In this case, the intervals would be the number of households with no child (0), one child (1), and two children (2) and so on.
- Make a table with separate columns for the interval numbers (the number of children per families), the tallied results, and the frequency of results in each interval. Label these columns as—number of children, tally and frequency (Table 5.2).
- Read the list of data from left to right and place a tally mark in the appropriate row. For example, the first result is 1, so place a tally mark in the row beside where 1 appears in the interval column (number of children). The next result is 2, so place a tally mark in the row beside the 2, and so on. When you reach your fifth tally mark, draw a tally line through the preceding four marks to make your final frequency calculations easier to read. Now it will look like a stack.
- Add up the number of tally marks in each row and record them in the final column entitled Frequency.

  By looking at this frequency distribution table we can quickly see that out of 20 families surveyed, 4 families had no children, 6 families had 1 and so on (Table 5.2).

TABLE 5.2: Frequency table for the number of children in each family

Number of children (x)	Tally	Frequency (f)
0	IIII	4
1	𝍸, I	6
2	𝍸	5
3	III	3
4	II	2
		Total = 20

Constructing a Cumulative Frequency Distribution Table

A cumulative frequency distribution table is more detailed table. It looks almost like a frequency distribution table, but it has a column with added values and it contains the cumulative frequency and there may be a column for cumulative percentage of the results, as well.

Class intervals:

- If a variable takes a large number of values, then it is comfortable to present and handle the data by grouping the values into class intervals. Continuous variables are presented in class intervals, while for discrete variables they can be grouped into class intervals or not. For example, we have set out age ranges for a study of young people, while allowing for the possibility that some older people may also fall into the scope of our study.
- The frequency of a class interval is the number of observations that occur in a particular predefined interval. For example, if 20 people aged 5–9 appear in our study data, the frequency in class 5–9 intervals will be 20.
- The endpoints of a class interval are the lowest and highest values that a variable can take.
- If we have data showing following class intervals—0–4 years, 5–9 years, 10–14 years, 15–19 years, 20– 24 years, and 25 years and over, then the endpoints of the first interval are 0 and 4 if the variables are discrete, and 0 and 4.999 if the variable is continuous. It means the values from 0–4 will fall under 0–4 years and so on.

Contd...

Class interval width:

- Class interval width is the difference between the lower endpoint of an interval and the lower endpoint of the next interval. Therefore, if our data has continuous intervals like 0–4, 5–9, etc., then the width of the first 5 intervals is 5, and the last interval is open, because no higher endpoint is assigned to it. The intervals could also be written as 0 – <5, 5 – <10, 10 – <15, 15 – <20, 20 – <25, and 25 and >25.
- For deciding on the width of the class intervals, you have to decide between having intervals short enough so that most of the observations fall in the same interval, but they should be long enough so that you do not end up with only one observation per interval. It is equally important to make sure that the class intervals are mutually exclusive.

Example: The ages of the participants in a survey were recorded as follows:

36, 48, 54, 92, 57, 63, 66, 76, 66, 80

Solution:

- Divide the results into intervals, and then count the number of results in each interval. Here, intervals of 10 will be appropriate. Again, 36 is the lowest age and 92 is the highest age, so start the intervals at 35–44 and end the intervals with 85–94.
- Make a table similar to the earlier frequency distribution table but with 3 extra columns.
  - **Lower value column:** Write the lower values of the result intervals in the first column. For example, in the first row, you will write the number 35.
  - **Upper value column:** Write the upper values of the result in the second column. For example, in the second row, you will write the number 44.
  - **Frequency:** Note down the number of times a result in the third column—a particular digit appears between the lower and upper values. For example, in the third row, you will write the number 1.
  - **Cumulative frequency:** Here, you will add the cumulative frequency. Because it is the first row, the cumulative frequency will remain the same. Whereas, in the second row, the frequency for the 35–44 interval (i.e., 1) is added to the frequency for the 45–54 interval (i.e., 2). Thus, the cumulative frequency will be 3 here, meaning we have 3 participants in the age group of 34–54.
  - **Percentage:** In this column, write the percentage of the frequency. To do so, divide each frequency by the total number of results and multiply by 100. Here, the frequency of the first row is 1 and the total number of results is 10 which will be equal to 10%.

    $10\ (1 \div 10) \times 100 = 10$
  - **Cumulative percentage:** In this column, divide the cumulative frequency by the total number of results, then to make a percentage, multiply by 100. Note that the last number in this column should be equal to 100.

    Now the cumulative frequency distribution table will look like this (Table 5.3):

TABLE 5.3: Ages of participants

Lower value	Upper value	Frequency (f)	Cumulative frequency	Percentage	Cumulative percentage
35	44	1	1	10	10
45	54	2	(2+1) = 3	20	30
55	64	2	(3+2) = 5	20	50
65	74	2	(5+2) = 7	20	70
75	84	2	(7+2) = 9	20	90
85	94	1	(9+1) = 10	10	100

Example: Construct a frequency distribution table for the large numbers of observations of diabetic patients.

423, 369, 387, 411, 393, 394, 371, 377, 389, 409, 392, 408, 431, 401, 363, 391, 405, 382, 400, 381, 399, 415, 428, 422, 396, 372, 410, 419, 386, 390

Solution:

After observing the data, we find that here the lowest value is 363 and the highest is 431. We will take a class interval of 10, the interval for the first class is 360–369 and includes 363 (the lowest value). Remember, there should always be enough class intervals so that the highest value is included. The completed frequency distribution table will look like this (Table 5.4):

TABLE 5.4: Sugar levels in diabetic patients

Classes (x)	360–369	370–379	380–389	390–399	400–409	410–419	420–429	430–439	Total
Tally marks	II	III	𝍸	𝍸 II	𝍸	IIII	III	I	
Frequency (f)	2	3	5	7	5	4	3	1	30

Practical Tips

Rules for data sets that contain very large number of observations

In short follow these basic rules while constructing a frequency distribution table for a data set which contains a large number of observations:

1. Find the lowest and highest values of the variables.
2. Decide on the width of the class intervals.
3. Include all possible values of the variable.

Relative Frequency and Percentage Frequency

Relative frequency and percentage frequency may also be required by a researcher, as he/she may like to know what proportion of the values falls into each class interval.

The relative frequency of a particular observation or class interval is calculated by dividing the frequency (**f**) by the number of observations (**n**): that is, (**f** ÷ **n**). Thus:

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Number of observations}}$$

Further the percentage frequency is calculated by multiplying each relative frequency value by 100.

Thus:

$$\text{Percentage frequency} = \text{relative frequency} \times 100$$

Example: Constructing relative frequency and percentage frequency in a table.

Solution:

Using the data of example 4, here is table of relative frequency and percent frequency:

After looking at the above data we can conclude that:

- 7% of values fall in the class 360–369 (in the first class for example).
- And the probability of any randomly selected observation in this range is approximately 0.07.
- The first column in the table here represents the marks obtained in class interval form. The lowest number in a class interval is called the **lower limit** and the highest number is called the **upper limit**. This example is the case of continuous class intervals as the upper limit of one class is the lower limit of the following class (Table 5.5).

TABLE 5.5: **Relative frequency and percentage frequency table**

Classes (x)	Frequency (f)	Relative frequency	Percent frequency
360–369	2	0.07	7
370–379	3	0.10	10
380–389	5	0.17	17
390–399	7	0.23	23
400–409	5	0.17	17
410–419	4	0.13	13
420–429	3	0.10	10
430–439	1	0.03	3
Total	**30**	**1.00**	**100**

Practical Tips

In continuous cases, any observation corresponding to the extreme values of a class is always included in that class where it is the lower limit. For example, if we had a student who has scored 5 marks in the test, his marks would be included in the class interval 5–10 and not 0–5. Here, we have assumed that a representative sample has been drawn but in the real world, researcher has to refer to an estimate of variability, in order to complete the analysis.

Example: The following is the distribution for the age of the students in a school:

Age	0–5	5–10	10–15	15–20
Number of students	35	45	50	30

Calculate the following:

- The lower limit of the first-class interval.
- The class limits of the third class.
- The class mark for the interval 5–10.
- The class size.

Solution:

- The lower limit of the first class interval, i.e., 0–5 is '0'.
- The class limits of the third class, i.e., 10–15 are 10 (lower limit) and 15 (upper limit).
- The class mark is defined as the average of the upper and the lower limits of a class. For 5–10, the class mark is $\frac{5+10}{2} = 7.5$.
- The class size is the difference between the lower and the upper class-limits. Here, we have a uniform class size, which is equal to 5 (5–0, 10–5, 15–10, 20–15, all are equal to 5).

STUDENT ASSIGNMENT

LONG ANSWER QUESTIONS

1. What is frequency distribution table? Discuss in detail.
2. Explain the relative frequency and percentage frequency with an example.

SHORT ANSWER QUESTIONS

1. What is a relative frequency?
2. Define frequency.
3. Define range.
4. Write a short note about the class interval.
5. What is an ungrouped data?
6. Define grouped data.

MULTIPLE CHOICE QUESTIONS

1. Type of cumulative frequency distribution in which class intervals are added in top to bottom order is classified as:

a. Variation distribution
b. Less than type distribution
c. More than type distribution
d. Marginal distribution

2. Type of cumulative frequency distribution in which class intervals are added in bottom to top order is classified as:

a. More than type distribution
b. Marginal distribution
c. Variation distribution
d. Less than type distribution

3. 'Less than type distribution' and 'more than type distribution' are types of:

a. Class distribution
b. Cumulative class distribution
c. Cumulative frequency distribution
d. Upper limit distribution

4. Distribution which shows cumulative figure of all observations placed below upper limit of classes in distribution is considered:

a. Cumulative frequency distribution
b. Upper limit distribution
c. Class distribution
d. Cumulative class distribution

5. Class frequency is divided by number of observations in frequency distribution to convert it into:

a. Relative margin distribution
b. Relative variable distribution
c. Relative frequency distribution
d. Relative width distribution

ANSWER KEY

1. b **2.** b **3.** c **4.** a **5.** c

Simplified

Biostatistics For Nurses

Salient Features

- The text is enriched with a variety of formulas, ensuring a deeper practical understanding of statistical analysis when applied to real-world scenarios.
- Simplified solutions are provided in the form of solved examples, making it easier to grasp the complex concepts and reinforcing understanding of each respective topic.
- Divided into 10 Units and 18 chapters, the book covers a wide range of biostatistical concepts—from basic principles to more advanced, complex topics—offering a thorough exploration of the subject.
- Numerous practical examples have been included with step-by-step solutions to illustrate the application of statistical procedures in real-time research and data analysis.
- Practical Tips boxes throughout the book, provide valuable insights and actionable advice, helping the students in the practical implementation of statistical methods.
- Must Know boxes with valuable facts are strategically placed to highlight critical information, ensuring readers are well-informed of key concepts and important details.

Learning Objectives enlist what the students will learn after studying the entire chapter.

LEARNING OBJECTIVES

After the completion of the chapter, the readers will be able to:

- Understand concepts of central tendencies.
- Know about the relation between the measures of central tendencies.

Chapter Outline provides a quick glance of the entire chapter in one go.

CHAPTER OUTLINE

- Introduction
- Central Tendency
- Mean
- Mode

Must Know boxes covering valuable facts are strategically placed to highlight critical information, ensuring readers are well-informed of key concepts and important details.

Must Know

Modifying a distribution by dumping scores or by addition of new scores will generally change the value of the mean and it will affect: Number of scores; Sum of the scores. If a constant value is added to every score in a distribution, then the same constant value is added to the mean. Also, if every score is multiplied by a constant value, then the mean is also multiplied by the same constant value.

Practical Tips boxes throughout the book, provide valuable insights and actionable advice, helping the students in the practical implementation of statistical methods.

Practical Tips

- In both the above given formulae '$n - 1$' is used instead of n in the denominator, because it gives a more accurate estimate of population SD.

Illustrations and Tables are used to make learning easy for students.

Figure 2.1: The wholesome picture of statistic and parameter

Table 2.1: Comparison of statistic and parameter

Characteristics	Statistic	Parameter
Definition	A characteristic of a small part of the population, i.e., sample.	A fixed measure that describes the target population.
Nature	A variable and known number that depend on the sample of the population.	Parameter is fixed and unknown numerical value.

Important and summarized facts of respective topic are covered under **Takeaway** boxes

Takeaway

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Total number of values}}$$

Median = Middle value (when the data are arranged in order)

Mode = Most common value

- Central tendency: A score which indicates a position where the center of a distribution tends to be located
- Mean is sum of all scores divided by the number of items

Detailed **Student Assignment** in the form of exercises in each and every chapter will facilitate structured learning and revision of the material provided in the respective chapters.

STUDENT ASSIGNMENT

LONG ANSWER QUESTIONS

1. What is the significance and scope of statistics?
2. How does statistics help in epidemiological studies?

SHORT ANSWER QUESTIONS

1. Write a short note on variables.
2. What are the applications of statistics in medical field?

About the Author

Anju Dhir, *PhD Microbiology,* is presently working as Senior Product Manager and Developmental Editor in Health Sciences division. She is a former Lecturer, Department of Microbiology at Shivalik Institute of Nursing, Shimla, Himachal Pradesh. She is a Gold Medalist in Microbiology. She had been in teaching profession for the last 25 years. Her thesis and research papers are published in national and international journals.

CBSPD
Dedicated to Education

CBS Publishers & Distributors Pvt. Ltd.

4819/XI, Prahlad Street, 24 Ansari Road, Daryaganj, New Delhi 110 002, India

E-mail: feedback@cbspd.com, **Website:** www.cbspd.com

New Delhi | Bengaluru | Chennai | Kochi | Kolkata | Lucknow | Mumbai
Hyderabad | Jharkhand | Nagpur | Patna | Pune | Uttarakhand

Scan QR Code to access CBS Nursing Catalogue

ISBN: 978-93-90619-47-4