**Cumulative frequency distribution**

ABCD EFGH

Department of , University Of Sargodha

CHAPTER NUMBER: CHAPTER NAME

Instructor name

DATE

**Contents**

Here’s a quick rundown of the contents of this guide on what is cumulative frequency distribution

1. Introduction 3
2. Brief overview 3
3. Purposes of frequency distribution 3
4. Advantages and disadvantages 4
5. Types of Cumulative frequency distribution 5
6. Cumulative frequency table 6
7. Characteristics of frequency distribution 7
8. Graphical Representation 8

9. **Cumulative frequency distribution as a data science tool 9**

**10. Numerical problems 10**

11. References 12

**INTRODUCTION**

. A cumulative frequency distribution is basically the sum of the class and all the classes below it in a frequency distribution. All that means is that you are adding up a value and all of the values that came before it. Cumulative Frequency distribution is a very common method used to organize Data. Frequency distribution basically an organized graphical representation of the number of individuals in each category on the scale of a measurements. It also gives easily access to the entire data. It shows whether the observations are high or low or remains constant across the entire scale of a data. Thus this shows that frequency distribution presents a picture of how the individual observe

The cumulative frequency distribution is also defined as the sum of all the frequencies up to and including that class. Cumulative frequency distribution is corresponding to the value of particular type of the variable, we either mean the number of observations is less than or equal to (\leq ) that value or the number of observations greater than or equal to (\geq ) that value is called the cumulative frequency of less than type and the cumulative frequency of greater than type respectively.

Specification of the different values of the variable together with the corresponding cumulative frequencies is called a cumulative frequency distribution which is generally represented in the form of a table where the values of the variable are written in order.

Cumulative Relative Frequency Distribution is the same as Cumulative Frequency distribution; just the difference is that in case of Cumulative Relative Frequency we add the relative frequencies instead of a simple frequencies. So from here we can conclude that in cumulative relative frequency distribution we can simply add up the relative frequencies while in cumulative frequency distribution we can add simple frequencies to obtained frequency distribution.

**Example:**

You get paid $100 for a week of work. The second week you get paid $200 and the third week, $300. Your cumulative amount for week 2 is $300 ($200 for week 2 and $100 for week 1). Your cumulative amount for week 3 is $600 ($300 for week 3, $200 for week 2 and $100 for week 1).

**PURPOSES OF CUMULATIVE FREQUENCY DISTRIBUTION:**

The main purpose of a cumulative frequency distribution is defined below:

1. The first main purpose of cumulative frequency distribution is to find the number of observations or proportion or percentage of observations less/more than a certain value or in between two given values or in between a given set of data.
2. The second main purpose of cumulative frequency distribution is to helps us to find values of different measures like medians, modes e

**ADVANTAGES AND DISADVANTAGES**

There are many advantages and disadvantages of cumulative frequency distribution. Some are listed below

**ADVANTAGES**

* Cumulative frequency distribution is the very useful way to determine the number of scores that occur until a certain value.
* There is also an original info from a grouped frequency distribution can be obtained from the CF curves.
* It will also show you whether the values are at constant rate or if they speeds up or slows down.
* Cumulative Frequency distribution tables can also be helpful in identifying the obvious trends within a data set and can also be used to compare the data between data sets of the same type.
* They are also more helpful in displaying descriptive statistics.
* They are also very easy to interpret and they can display large data sets in a fairly concise manner
* Cumulative Frequency table can also help researchers to audit the relative abundance of each particular target data group within their sample.
* Cumulative Frequency distribution tables can quickly acknowledge outliers and even compelling trends within a data set with not much more than a cursory inspection.

**DISADVANTAGES**

* Sometimes there is also very difficult to compare the frequencies between each data group
* It also requires high bandwidth
* In cumulative frequency distribution it is also sometimes difficult to compare different data sets

**TYPES OF CUMULATIVE FREQUENCY CURVES**

**There are two types of cumulative frequency distribution which are listed below**

* More than cumulative frequency Distribution
* Less than cumulative frequency distribution

### **1- MORE THAN CUMULATIVE FREQUENCY DISTRIBUTION**

The Number of observations greater than or equal to lower boundary of a class is called "more than type" cumulative frequency of that class. Similarly, cumulative frequency distribution of greater than type for a value of particular type of the variable is obtained by adding the frequencies of all values greater than that value, starting from last to the first, i.e., by adding its frequency to the frequencies of all the values greater than that value.

### 

**2- LESS THAN CUMULATIVE FREQUENCY DISTRIBUTION**

The Number of observations less than the upper boundary of a class is called "less than type" cumulative frequency of that class. Thus, cumulative frequency distribution of less than type for a value of particular type of the variable is obtained by adding the frequencies of all values less than that value up to the last frequency, i.e., by adding its frequency to the frequencies of all the values smaller than that value.

**CUMMULATIVE FREQUENCY TABLE**

A cumulative frequency distribution table shows the different measurement categories and the number of observations in each category. Before constructing a frequency table, one should have an idea about the range (minimum and maximum values). On the other hand, if they are very few, then the shape of the distribution itself cannot be determined. The width of the class can be determined by dividing the range of observations by the number of classes. The following are some guidelines regarding class widths:

• It is advisable to have equal class widths. Unequal class widths should be used only when large gaps exist in data.

• The class intervals should be mutually exclusive and no overlapping.

• Open-ended classes at the lower and upper side (e.g., 100) should be avoided.

The frequency distribution table of the resting pulse rate in healthy individuals is given in Table1.

**TABLE 1**

*Pulse rate of a healthy human being*

|  |  |  |  |
| --- | --- | --- | --- |
| **Pulse/min** | **Frequency** | **Pulse/min** | **Cumulative frequency** |
| 60-64 | 2 | Less than 64 | 2 |
| 65-69 | 7 | Less than 69 | 9 |
| 70-74 | 11 | Less than 74 | 20 |
| 75-79 | 15 | Less than 79 | 35 |
| 80-84 | 10 | Less than 84 | 45 |
| 85-89 | 9 | Less than 89 | 54 |
| 90-94 | 6 | Less than 94 | 60 |

Note: This table is for less than cumulative frequency distribution type.

**TABLE 2**

*Pulse rate of a healthy human being*

|  |  |  |  |
| --- | --- | --- | --- |
| **Pulse/min** | **Frequency** | **Pulse/min** | **Cumulative frequency** |
| 60-64 | 2 | More than 60 | 60 |
| 65-69 | 7 | More than 65 | 58 |
| 70-74 | 11 | More than 70 | 51 |
| 75-79 | 15 | More than 75 | 40 |
| 80-84 | 10 | More than 80 | 25 |
| 85-89 | 9 | More than 85 | 15 |
| 90-94 | 6 | More than 90 | 6 |

Note: This table is for more than cumulative frequency distribution type.

**CHARACTERISTICS OF FREQUENCY DISTRIBUTION**

There are four very important characteristics of frequency distribution

• Measures of mean, median, mode.

• Measures of range, variance, and standard deviation. These are also dispersions.

• The extent of skewness or symmetry/asymmetry

• The kurtosis which is also defined as flatness or peakedness.

**GRAPHICAL REPRESENTATION**

A frequency distribution graph is a diagrammatic illustration of the information in the frequency table.

Group of a data can also be represented by a curve called ogive or cumulative-frequency curve. As the name suggests, in this representation cumulative frequencies of different class intervals play an important role.

Ogives are the graphs that are used to estimate that how many numbers lies above or below a particular variable or value in data. To draw an Ogive graph, firstly, the cumulative frequency distribution of the variables is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value of the cumulative frequency will always be equal to the total of all observations.

The most commonly used graphs for the frequency distribution are histogram, [frequency polygon](https://byjus.com/maths/frequency-polygons/), frequency curve, and Ogive (cumulative frequency curves). Representation of cumulative frequency graphically is easy and convenient as compared to representing it using table, bar-graph, frequency polygon etc.

**HISTOGRAM:**

A histogram is basically a graphical representation of the variable in the *X* axis and the number of observations (frequency) in the *Y* axis. A histogram is used when we want to depict the frequency of a particular data which is measured. [Figure 1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117575/figure/F0001/) depicts a histogram constructed for the data given in [Table 1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117575/table/T0001/).

[[An external file that holds a picture, illustration, etc.
Object name is JPP-2-54-g001.jpg](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117575/figure/F0001/)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117575/figure/F0001/" \t "figure)

## 

## ****CUMULATIVE FREQUENCY DISTRIBUTION AS A DATA SCIENCE TOOL****

Cumulative frequency distribution is basically plays a very important role in the analysis of a big data, as it plays a very key role in extracting information that is used for data analysis purposes, such as decision making, developments of product, trend analysis and forecasting.

You should consider an example that data scientists plays a very important role in calculating the number of people which are suffering from heart diseases under certain weight category. After again very careful examination, scientists arrive at a conclusion that people having higher obesity has higher chances of suffering from a heart disease.

Now the question arises is that How did they arrive at this conclusion? They only arrive at this conclusion by observing the cumulative frequency distribution tables and graphs. Similar to what we observed in the upper table and figure examples. They compared the normal human being pulse rate with the minutes and made a table and graph from which they deduce the result. So we can conclude that cumulative frequency distribution can play a very vital role in analysis of a data.

**NUMERICAL PROBLEM**

The marks of 150 students in a test were recorded and shown by the following frequency distribution.

|  |  |
| --- | --- |
| **Marks %**  10 - 19  20 - 29  30 - 39  40 - 49  50 - 59  60 - 69  70 - 79  80 - 89 | **Number of Students**  7  11  10  26  37  37  15  7 |

Construct the cumulative frequency table.

Also answer the following.

(i) How many students obtained less than 40 marks?

(ii) How many students obtained at least 70 marks?

**SOLUTION**

The cumulative frequency table is as given below.

**Class Interval Frequency Cumulative Frequency**

10 – 19 7 7

20 – 29 11 18

30 – 39 10 28

40 – 49 26 54

50 – 59 37 91

60 – 69 37 128

70 – 79 15 143

80 – 89 7 150

1. The number of students which are obtaining less than 40 marks are

       = the cumulative frequency of the class interval 30 - 39 = 28.

Or

= Sum of the frequencies of the class interval 10-19, 20-29, and 30-39.

= 7+11+10

= 28.

2. The number of students obtaining at least 70 marks

    = Sum of the frequencies of the class intervals 70 - 79 and 80 - 89

      = 15 + 7

      = 22.

**REFERENCES:**

Figure 1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117575/figure/F0001

1. Gravetter FJ, Wallnau LB. Statistics for the behavioral sciences. 5th ed. Belmont: Wadsworth – Thomson Learning; 2000

2. Dawson B, Trapp RG. Basic and clinical biostatistics. 4th Ed. New York: McGraw Hill; 2004.

3. Sundaram KR, Dwivedi SN, Sreenivas V. Medical statistics principles and methods. 1st Ed. New Delhi: B.I Publications Pvt Ltd; 2010.

4. Swinscow TDV, Campbell MJ. Statistics at square one. 10th Ed (Indian). New Delhi: Viva Books Private Limited; 2003.

5. Norman GR, Streiner DL. Biostatistics the bare essentials. 2nd ed. Hamilton: B.C. Decker Inc.; 2000.

6. Sundar Rao PS, Richard J. Introduction to biostatistics and research methods. 4th Ed. New Delhi: Prentice Hall of India Pvt Ltd; 2006.