Statistics

DPT

AWAIS HAMZA

Representation Of Data

- To extract information from the sample, there is need to organize and summarize the collected data.
- There are many ways of describing a sample. Commonly, we use either, a graph or a small set of numbers which summarize some properties of the sample such as its centre and spread.

Classification

- The term classification is the process of arranging observations into different classes or categories according to some common characteristics.

Classification

- The data may be classified by one or more characteristics at a time.When data is classified according to one characteristic, it is called **one-way classification.**
  Example:
  Students classified **by gender** – boys and girls.

- When the data is classified by two characteristics at a time, it is called **two-way classification.**
- **Example:**
  Students classified **by gender** – boys and girls.

Classification

- Similarly, the data classified by three characteristics is called **three-way classification.**
- **Example:**
  Students classified **by gender and grade** – boys and girls in Grade A, B, and C.

Tabulation

- The process of making tables or arranging data into rows and columns is called tabulation.
- Tabulation may be **simple, double, triple or complex** depending upon the number of characteristics involved.
- Tables are the most common form of documentation used by the scientists.

Construction of tables

- **Title:** A title is the **heading at top of the table**. The title should be brief and self explanatory. It describes the contents of the table.
- **Column Captions and Box head**: The headings for different columns are called **column captions** and this part of column captions is called **Box-head**.

Construction of tables

- **Row Captions and Stub**: The headings for different rows are called **row captions** and the part of the table containing row captions is called **Stub.**
- **Body of the Table and Arrangements of the data:** The entries in different cells of column and rows in a table are called body of the table.
- It is the main part of the table.
- **Source Note:** Source notes are given at the end of the table which indicate the compiling agency, publication, the data and page of the publication.

Construction of tables

- **Spacing and ruling**: To enhance the effectiveness of a table, spacing and ruling is used.
- It is also used to separate certain items in the table.
- Thick or double lines or single lines are used to separate row captions and column captions.
- To indicate no entry in a cell of the table, dots (...) or dashes(---) are used. Zeroes are not used in a table for this purpose.

Construction of tables

- **Prefatory Notes and Footnotes:** The prefatory note is given after the title of the table and the footnotes are given at the bottom of the table. Both are used to explain the contents of the table. The footnotes are usually indicated by***

Table Number: .....
Title: .....
(Head Note, if any)

Stub (Row Heading)	Caption (Column Heading)				Total (Rows)
	Sub-head		Sub-head		
	Column-head	Column-head	Column-head	Column-head	
Stub Entries (Row Entries)		Body			
Total Columns					

Source Note:
Footnote:

Frequency Distribution

- . A **frequency distribution** is a compact form of data in a table which displays the categories of observations according to their magnitudes and frequencies such that the similar or identical numerical values are grouped together.
- The categories arc also known as groups, class intervals or classes.

Frequency Distribution

- The relative frequency, denoted by r.f of a category is the proportion of observed frequency to the total frequency and is obtained by dividing observed frequency by the total frequency.
- The sum of the relative frequencies should be one (1) except for rounding error. The relative frequencies are important for making comparisons between two or more distributions. Otherwise, the different sample sizes of the data sets may distort comparisons.

Frequency Distribution

- Following steps are taken into account while making a frequency distribution for continuous data:

1. Calculate range of the data, where Range = maximum value in the data. - minimum value in the data
2. Decide about the number of classes. The minimum number of classes may be determined by the formula

   Number of classes c = 1+3.3 log(n) or c=√n (approximately)

   where n is the total number of observations in the data.

Frequency Distribution

- Decide about width of the class. It is usually abbreviated by h and is obtained by the following relation:
- **h = Range/number of classes**
- It should be noted that always a convenient near number is chosen and it is not necessary to follow the rules of rounding because we are only grouping the data.

Frequency Distribution

- The decision about the starting point of the first class is arbitrary usually, it is started before the minimum value in such a way that the mid point, the average of lower and upper class limits of the first class is properly placed.

- Now, an observation is taken and a mark of vertical bar is made for a class it belongs. A running tally is kept till the last observation. The tally count N indicates five.

Example

- Student- Height Data The height (in cms) of 30 students measured at the time of registration is given by 91,89,88, 87,89,91, 87, 92, 90, 98,95,97,96, 100, 101, 96, 98,99,98, 100, 102, 99, 101, 105, 103, 107, 105, 106, 107, 112. **Make a suitable frequency distribution.**

Solution: Step 1

- Range=Maximum value - minimum value
- In this data maximum value is 112 and minimum value is 87.
- So, Range=112 - 87=25

Step 2

- Approximate number of classes or class intervals are number of classes c is given by
- = 1 + 3.3 log(30) =1+3.3 (1.4771) = 5.87443 = 6 (approximately)

Step 3

- Width of the class interval(h)
- **h = range / number of classes**
- 25/6 =4.167 = 5 (approximately)
- 5 is chosen for convenience, one may take 4 if he / she wishes so.

Step 4

- Minimum value is 87, we start the first class from 86 with width of the class as 5, so, our first class is 86-90 with mid point 88, the average of lower and upper class limits i.e., (86+90)/2=88.
- Similarly, other classes are 91-95, 96-100,...., 111-115.
- It is clear that maximum value 112 is included in the last class.

Step 4

- It is convenient to choose the end points of the class interval so that no observation falls on them.
- This can be obtained by expressing the end points to one more place of decimal than the observations themselves.
- Therefore, suitable class boundaries for this data would be 85.5-90.5, 90.5-95.5, …….110.5 115.5. In the class boundaries, the upper values in the classes are included in the next class so that the classes are mutually exclusive i.e., 90.5 is the upper value of the first class and is lower value of the second class. counting this would be included in the second class interval.

Step 4

- The class centres Y's are the middles of the classes. T
- he class centres are also known as mid or middle points and are obtained either by averaging class limits or class boundaries i.e
- Y is the middle of the first class Y = (85.5+90.5)/2=88 v) The other mid points are 93,98,...., 113 respectively.

Step 5

- Starting from first observation, all the 30 observations are assigned to the classes they belong.
- The first observation 87 falls in the first class 86-90, a tally mark is made in the tally column against this class.
- The second observation 90 belongs to the first class 86-90, a tally mark is made in tally column against this class and so on, the last observation 112 belongs to the last class 111-115. The number of tally marks in the tally column against each class gives the frequency of that class.

Class Limits	**Class Boundaries**	**Midpoint (X)**	**Frequency (f)**
86 – 90	85.5 – 90.5	88	6
91 – 95	90.5 – 95.5	93	4
96 – 100	95.5 – 100.5	98	10
101 – 105	100.5 – 105.5	103	6
106 – 110	105.5 – 110.5	108	3
111 – 115	110.5 – 115.5	113	1
Total			**30**


Any
Question?


Thank You