UNIT 13

Statistical Techniques

Unit Overview

Learning Objectives

- *To Understand the concepts of geographical data*
- *To know the various types and methods of data collection.*
- *To analyse and interpret various data used in geography*
- *To develop and understand the measures of central tendency*

35LDO

13.1 Introduction

Statistics is a branch of science which deals with quantitative data. Actually statistics includes in four processes.

1. Collection of Data
2. Presentation of Data
3. Analysis of Data
4. Interpretation of Data

13.2 Role of statistics in geography

Statistical techniques and procedures are applied in all fields of academic research; wherever data are collected and summarized or wherever any numerical information is analyzed or research is conducted, statistics are needed for sound analysis and interpretation of results.

Geographers use statistics in numerous ways

- To describe and summarize spatial data.
- To make generalizations concerning complex spatial patterns.
- To estimate the probability of outcomes for an event at a given location.
- To use samples of geographic data to infer characteristics for a larger set of geographic data (population).
- To determine if the magnitude or frequency of some phenomenon differs from one location to another.
- To learn whether an actual spatial pattern matches some expected pattern.

For example, a student studying the annual rainfall in a particular region will first of all

collect the data regarding rainfall of the region. Then he will summarize the collected data in some table or diagram to show the desired results.

13.3 Types of Data

Data can be primarily divided into two types – qualitative data and quantitative data. Qualitative data only describes an item example coconut tree, palm tree, etc or language spoken – Tamil, Telugu, Hindi, English, etc. Quantitative data on the other hand has numerical value to the item, example, marks obtained - 45, 70, 60, 90, 25, etc.

Data can also be divided based on the source or method of collection as the primary data and secondary data. The primary data are those statistical materials which the investigator collects by going from door to door for the purpose of his/her study. On the other hand, the secondary data refer to already collected and recorded data which the investigator had collected from published or unpublished sources.

13.4 Methods of collection of Data

Methods of Primary data collection

1. Direct personal investigation (interview)
2. Indirect oral investigation
3. Questionnaire Method

Methods of Secondary data collection

1. Published Sources
2. Unpublished Sources

13.5 Organisation and Classification of Data

After collecting data it has to be classified and arranged for further statistical analysis. The data classified is large in quantity needs to be reduced for further analysis. This leads to organization of data.

Following steps are important in organization of data

Finding Range

While organizing the data, it is necessary to find out the highest and lowest number (maximum and minimum value) from the given data. The difference between these two extreme values is called range.

Range = Highest value – Lowest value.

Example

Find out the range from the following data 2, 6, 4, 9, 12, 8, 5, 8. In this case, highest value is 12 and lowest value is 2, and the range is 12 – 2 = 10.

Frequency array

In statistics, array refers to some kind of orderly arrangement. When the numerical raw data is arranged in ascending or descending order, it is called array.

Frequency distribution

It is an arrangement of data into classes. Following technical terms are important for frequency distribution.

1. Class

It is a group which divides the variable into parts and forms a set of given frequency.

Example

CLASS INTERVAL	TALLY	FREQUENCY
0-10	|	1
10-20	𝍸	5
20-30	𝍸 𝍸 ||	12
30-40	𝍸 |||	8
40-50	||||	4
	SUM	30

2. Class limit

The extreme values of the classes or groups are called the class limit. Each class has lower limit (minimum value) and upper limit (maximum value).

For example in the class 10 – 20, the lower limit is 10 and upper limit is 20.

3. Midpoint

It is obtained by dividing the sum of lower limit and upper limit by 2.

For example in the class 10 – 20, the midpoint is $\frac{10+20}{2}=15$

4. Class interval

It is the difference between the upper and lower limit of a class.

For example:

In the class 10 – 20,

The class interval is 20 – 10 = 10.

5. Tally Marks

Tally marks usually come in group of 5. For every observation, we put vertical bars called tally marks. After 4 bars we put a cross stroke for 5th observation. It helps in easy counting of the frequencies.

6. Class Frequency

The number of observations included in each class is the class frequency or sum of tally marks of that particular class.

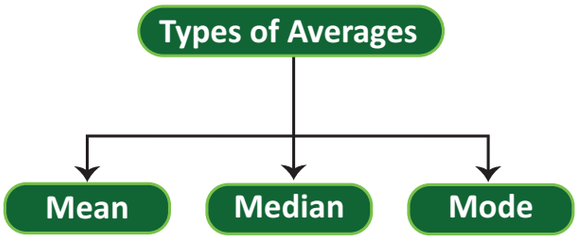
7. Total Frequency

It is an aggregate of all class frequencies. It is the sum of all observations.

13.6 Statistical techniques - Measures of central tendency

One of the most important objectives of statistical analysis is to get one single value that describes the characteristics of the entire data. A common word used for measures of central tendency is average. For example, if we collect the heights of 5000 students of a college, we will have 5000 figures. The persons mind is not capable of registering all the data at a time. So we need a single figure which represents the whole data. This single figure is known as average.

Since an average represents the entire data, its value lies somewhere in between the two extremes i.e. the largest observation and the smallest observation. For this reason the average is frequently referred as measure of central tendency.



Mean

Mean is defined as the value obtained by dividing sum of values of all the given items by the number of items. It is also called as arithmetic mean.

Formula

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

Mathematically it can be shown as

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots X_n}{N}$$

$$\bar{X} = \frac{\Sigma X}{N}$$

$\bar{X}$ = Mean

ΣX = Sum of the Values of all items

N = Total number of items

Calculation of Mean

Example 13.1

The following are the monthly average temperature data of Cuddalore District. Find the mean annual temperature.

Cuddalore

Temperature^0 C	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
	25.3	26.58	28.46	30.61	31.32	30.75	30	29.34	29.03	27.89	26.45	25.36

Solution:

Formula:

$$\bar{X} = \frac{\sum X}{N}$$

$$\bar{X} = \frac{25.3+26.58+28.46+30.61+31.32+30.75+30+29.34+29.03+27.89+26.45+25.36}{12}$$

$$\bar{X} = \frac{341.09}{12}$$

$$\bar{X} = 28.42$$

Answer:

The Mean annual temperature of Cuddalore district is 28.42

Calculation of Mean for Grouped Data or Continuous Series

Formula:

$$\bar{X} = A + \frac{\sum fd}{N} \times i$$

Where

A = Assumed mean
f = frequency
i = class size
d = deviation from assumed mean
N = Total

$$d = \frac{x - A}{i}$$

Example 13.2

Calculate the mean for the following data of heights of the plants in a garden

Heights in cms	10-20	20-30	30-40	40-50	50-60
No. of plants	5	15	10	15	5

Solution:

Heights in cms X	No of Plants F	Mid point x	A = 25, i = 10, $d = \frac{x - A}{i}$	fd
10 – 20	5	15	-1	-5
20 – 30	15	25 A	0	0
30 – 40	10	35	1	10
40 – 50	15	45	2	30
50 – 60	5	55	3	15
	N =50			**Σfd =55-5 = 50**

A = 25 , ∑ fd = 50 , N = 50, i = 10

Formula:

$$\bar{X} = A + \frac{\sum fd}{N} \times i$$

$$\bar{X} = 25 + \frac{50}{50} \times 10$$

$$\bar{X} = 25 + 10$$

$$\bar{X} = 35$$

Median:

Median is the middle value of a series arranged in any order – ascending or descending. Median divides the series in to two equal parts.

Calculation of Median

Example 13.3

The following are the monthly minimum temperature of Chennai District. Find the median.

Chennai

Minimum Temperature ° C					
Jan	21.2	May	28	Sep	25.6
Feb	22.2	June	27.5	Oct	24.6
Mar	24.2	July	26.4	Nov	23.1
April	26.6	Aug	25.9	Dec	21.9

Solution:

Step 1:

Arrange the data in Ascending Order

1	2	3	4	5	6	7	8	9	10	11	12
21.2	21.9	22.2	23.1	24.2	24.6	25.6	25.9	26.4	26.6	27.5	28

Step 2:

Formula

$$M = \text{Value of the } \left(\frac{N+1}{2}\right)^{th} \text{ item}$$

$$M = \text{Value of the } \left(\frac{12+1}{2}\right)^{th} \text{ item}$$

$$M = \text{Value of the } \left(\frac{13}{2}\right)^{th} \text{ item}$$

$$M = \text{Value of the } 6.5^{th} \text{ item}$$

$$M = \frac{\text{Size of } 6^{th} \text{ item + size of } 7^{th} \text{ item}}{2}$$

$$M = \frac{24.6 + 25.6}{2}$$

Answer: M = 25.1

Calculation of Median for Grouped Data or Continuous Series

Example 13.4

Calculate the median for the following data

Literates by age in 2001

Age group	Literacy in %
5 – 10	22.3
10 – 15	81.7
15 – 20	79.3
20 – 25	73.2
30 – 35	64.5
35 – 40	50

Solution:

Age group	Frequency f	Cumulative Frequency c f	
5 – 10	22.3	22.3	
10 – 15	81.7	104	
15 – 20	79.3	183.3	
20 – 25	**73.2**	**256.5**	→ Median class
30 – 35	64.5	321	
35 – 40	50	371	
	N = 371		

Calculation of Median Class

$$N = 371,\ \left(\frac{N}{2}\right) = \left(\frac{371}{2}\right) = 185.5$$

Class corresponding to 185.5 is 20-25 i.e. the Median Class

Formula

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

where

l = Lower limit of the median class
N = Total Numbers of frequeny
f = Frequeny of the median class
m = Cumulative frequency of the class preceding the median class
c = The class interval of the median class.

Formula

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

where $l = 20$, f $= 73.2$, m $= 183.3$, c $= 5$

$$\text{Median} = 20 + \left(\frac{185.5 - 183.3}{73.2}\right) \times 5$$

$$\text{Median} = 20 + \left(\frac{2.2}{73.2}\right) \times 5$$

$$\text{Median} = 20 + (0.03 \times 5)$$

$$\text{Median} = 20 + 0.15$$

Answer: Median = 20.15

Mode

Mode is that value which occurs most frequently in a set of observation or the item which repeats itself the greatest number of times

Calculation of Mode

Example 13.5

Calculate mode for the following data

Station: Cochin

Direction	% of days wind blowing
N	2
NE	10
E	10
SE	6
S	25
SW	25
W	5
NW	7
Calm	10

Solution:

From the data 25 is repeated 3 times. Hence the Mode is 25

Answer: Mode = 25

Calculation of mode for Grouped Data

Example 13.6

Calculate the mode for following data

Daily Wages	No of Workers
0 – 10	3
10 – 20	5
20 – 30	7
30 – 40	10
40 – 50	12
50 – 60	15
60 – 70	12
70 – 80	6
80 – 90	2
90 – 100	8

Solution

Daily Wages	No of Workers	
0 – 10	3	
10 – 20	5	
20 – 30	7	
30 – 40	10	
40 – 50	12 f_0	
50 – 60	15 f_1	→ Modal class
60 – 70	12 f_2	
70 – 80	6	
80 – 90	2	
90 – 100	8	

Highest Frequency = 15

Therefore 50 – 60 is the modal class

Formula

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

Modal class is the class which has maximum frequency.

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

c = width of the class limits

l = lower limit of mode interval

$l = 50,\ f_0 = 12,\ f_1 = 15, f_2 = 12$

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

$$\text{Mode} = 50 + \left(\frac{15-12}{2\times15-12-12}\right)\times10$$

$$\text{Mode} = 50 + \left(\frac{3}{30-24}\right)\times10$$

$$\text{Mode} = 50 + \left(\frac{30}{6}\right)$$

$$\text{Mode} = 50 + 5$$

Answer: **Mode = 55**

Exercise

1. The following are the monthly average rainfall of Chennai district. Find the mean, median and mode.

Chennai

Rainfall in mm	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
	15.54	10.44	12.48	20.96	57.1	41.24	64.4	99.28	147.17	204.22	165.07	133.76

2. Calculate the mean, median and mode for the following data relating to the monthly income of workers in a factory

Income in thousands	15–25	25–35	35–45	45–55	55–65	65–75	75–85
No of persons	6	11	7	4	4	2	1

3. Calculate mean, median and mode for the following data.

Annual rainfall in cms	0–25	25–50	50–75	75–100	100–125	125–150	150–175	175–200
Area in Sq.m	300	460	650	730	1200	1900	1500	660

References

1. Elements of Practical Geography, R.L. Singh and Rana P.B. Singh.
2. Advanced Practical Geography, Piyushkanti Saha and Partha Basu.