GEOG-510

QUANTITATIVE TECHNIQUES IN GEOGRAPHY

DEPARTMENT OF GEOGRAPHY AND NATURAL RESOURCE MANAGEMENT

SCHOOL OF EARTH AND ENVIRONMENTAL SCIENCE
UTTARAKHAND OPEN UNIVERSITY

(Teenpani Bypass, Behind Transport Nagar Haldwani (Nainital) Uttarakhand)

M.A /M.Sc GEOG - 510

QUANTITATIVE TECHNIQUES IN GEOGRAPHY

DEPARTMENT OF GEOGRAPHY

AND NATURAL RESOURCE MANAGEMENT

SCHOOL OF EARTH AND ENVIRONMENTAL SCIENCE

UTTARAKHAND OPEN UNIVERSITY

Phone No. 05946-261122, 261123

Toll free no. 18001804025

Fax no. 05946-164232, E. Mail info@uou.ac.in

Website: https://uou.ac.in

Board of Studies

Programme Coordinator

Dr. Ranju Joshi Pandey

Department of Geography and Natural Resource Management

School of Earth and Environmental Science

Uttarakhand Open University, Haldwani

Unit Writers

S.No.	Unit Written By	Unit No.
1.	Dr. Mohan Singh Sammal Assistant Professor School of Geography& NRM Uttarakhand Open University, Haldwani	1 & 4
2.	Mr. Sudhanshu Kumar Verma Assistant Professor School of Geography& NRM Uttarakhand Open University, Haldwani	2
3.	Mr. Sunil Tewari Assistant Professor School of Geography& NRM Uttarakhand Open University, Haldwani	3

EDITORS

Title	: Quantitative Techniques in Geography
ISBN	:
Copyright	: Uttarakhand Open University
Edition	: First (2024)

Published By : Uttarakhand Open University, Haldwani, Nainital-263139

Printed By :

CONTENTS

CONTENTS		PAGE NO.
BLOCK- 1 STATISTICAL METHODS		
UNIT-1	Measure of dispersion: variability, range, mean deviation, quartile deviation, standard deviation.	**1-23**
UNIT-2	Sampling theory: Basic concepts of probability, test & significance ‘t’ test & ‘chi square’ test , cental tendency, mean, median, mode, central tendency.	**24-46**
UNIT-3	Correlation: Karl Pearson & Spearman’s Regression analyses Cartography & representation of socio-economic data: dots, proportionan squires, circles, spheres, coropleth, star & scatter diagrams, Ergographs, flow line diagrams.	**47-84**
UNIT-4	Contouring and Determination of Heights with Indian Pattern Clinometer.	**85-103**

BLOCK- 1 STATISTICAL METHODS

UNIT-1 MEASURES OF DISPERSION: VARIABILITY, RANGE, MEAN DEVIATION, QUARTILE DEVIATION AND STANDARD DEVIATION

1.1 OBJECTIVES
1.2 INTRODUCTION
1.3 MEASURES OF DISPERSION: VARIABILITY, RANGE, MEAN DEVIATION, QUARTILE DEVIATION AND STANDARD DEVIATION
1.4 SUMMARY
1.5 GLOSSARY
1.6 ANSWER TO CHECK YOUR PROGRESS
1.7 REFERENCES
1.8 TERMINAL QUESTIONS

1.1 OBJECTIVES

After reading this unit the learners should be able to aware the following objectives:

- To understand the concept of dispersion.
- To know the limitations of average.
- To know the different measures of dispersion.
- To understand the importance of dispersion measurements in practical.

1.2 INTRODUCTION

In the world of data analysis, numbers often tell stories. They reveal trends, patterns, and insights that help us understand the world around us. But numbers, as powerful as they are, sometimes need a little extra explanation to truly understand the whole picture. This is where measures of dispersion come into play.

When we discuss about dispersion a question rise in our mind that what is Dispersion? Dispersion essentially tells you how much the individual data points deviate from the central value, typically the mean or median. It's like looking beyond the average to see how the numbers are behaving on their own. Are they huddled close together or scattered far apart? Dispersion measures, also known as measures of variability, quantify the extent to which individual data points in a dataset deviate or spread out from a central value, such as the mean or median. These measures provide insights into the distribution and spread of data points, allowing for a deeper understanding of the variability within a dataset.

Understanding dispersion matters for several reasons. It gives you insights into the reliability of your data – the less dispersion, the more consistent and dependable your measurements. Dispersion can also signal important characteristics about the data's distribution. For instance, a narrow dispersion might suggest a more concentrated distribution, while a wider dispersion could imply a broader spread. Measures of dispersion help us move beyond the averages and understand the full story within our data. Whether it's the simplicity of the range, the robustness of the quartile range, or the precision of the standard deviation, each measure provides a different angle on how our numbers are behaving. By grasping these measures, we gain a richer understanding of the patterns and behaviors underlying our data.

For example, imagine you have a group of friends and you are collecting their test scores. One friend scores 90, another 85, and yet another scores 95. These numbers give you an idea of the average performance, but do they tell you everything? Not quite. Measures of dispersion fill in the gaps by showing you how spread out or clustered these numbers are.

In brief, measures of dispersion bring balance to our understanding of data. They help us move beyond mere averages and allow us to see the dynamics, fluctuations, and uniqueness that individual data points can bring to the table. By appreciating these measures, we can make more accurate interpretations, predictions, and decisions based on the rich tapestry of information woven within our data.

1.3 MEASURES OF DISPERSION: VARIABILITY, RANGE, QUARTILE DEVIATION, MEAN DEVIATION AND STANDARD DEVIATION

As discussed in introduction section, the degree to which values in a distribution deviate from the distribution's average is known as the dispersion. To examine the level of the variation, there are following measures namely:

1. Variability
2. Range
3. Quartile deviation
4. Mean deviation and
5. Standard deviation

For deeper understanding, a detailed account of these dispersion measures is presented in the following paragraphs.

1. Variability

Measures of dispersion are also known as measures of variability. Variability refers to the extent to which individual data points in a dataset deviate or spread out from a central measure of tendency, such as the mean or median. In other words, it quantifies how much the data points vary or differ from the typical or central value of the dataset. Measures of dispersion provide insight into the spread and distribution of data, helping us understand the range and distribution of values within the dataset.

Greater variability indicates that the data points are more spread out and diverse, while lower variability suggests that the data points are closer to the central value and exhibit less scatter. Several measures of dispersion, such as the range, quartile range (QR) and standard deviation, help quantify and describe this variability. These measures provide different perspectives on how the data points are distributed around the central tendency.

2. Range

The range is one of the simplest measures of dispersion. It quantifies the spread of data in a dataset by calculating the difference between the largest (maximum) and smallest (minimum) values. In essence, it provides the interval over which the data points vary while the range is easy to calculate and understand. While easy to compute, the range is sensitive to outliers and does not capture the full distribution of data points.

Formula:

R= L-S

Where, R= Range, L= Largest and S= Smallest.

Example: Let's say we have a dataset representing the ages of a group of individuals:

20, 25, 22, 28, 30, 32, 19, 23, 27, 40

To calculate the range:

Find the largest value= 40

Find the smallest value= 19

Calculate the range: 40 – 19= 21

In this example, the range is 21, which means the ages of the individuals vary by 21 years, from the youngest (19 years old) to the oldest (40 years old). However, the range doesn't provide information about the distribution of ages between these two extremes.

3. Quartile Deviation

The usefulness of range as a measure of dispersion can be diminished by the existence of even one extraordinarily high or low number in a distribution. As a result, you might need a measurement that is not overly impacted by outliers. The quartile deviation (also known as semi-inter quartile range or semi-IQR) is a measure of dispersion that quantifies the spread of data in a dataset based on the inter quartile range (IQR). The IQR is the range between the first quartile (Q1) and the third quartile (Q3) of the dataset, and it contains the middle 50% of the data. The quartile deviation, being half of the IQR, gives a measure of the dispersion of the central half of the data. The quartile deviation is less sensitive to extreme values compared to the range, making it a more robust measure of dispersion. It provides insight into the spread of the central portion of the data distribution.

Formula:

Q.D. = $Q_3 – Q_1/2$

Where= Q.D. = Quartile deviation, Q_3= Third quartile and Q_1= First quartile.

Example: Let's say we have a dataset representing the ages of a group of individuals:

20, 25, 22, 28, 30, 32, 19, 23, 27, 40

To calculate the quartile deviation:

1. Calculate the first quartile (Q_1) and third quartile (Q_3) of the dataset.
   First, arrange the data in ascending order: 19, 20, 22, 23, 25, 27, 28, 30, 32, and 40.
   Q_1 corresponds to the median of the lower half of the data: Q_1: 20+22/2 = 21
   Q_3 corresponds to the median of the upper half of the data: Q_3: 30+32/2 = 31
2. Calculate the quartile deviation: Q.D.= 31 – 21/2 = 5

In this example, the quartile deviation is 5, indicating that the middle 50% of the ages in the dataset vary by approximately 5 years. The quartile deviation provides insight into the spread of the central portion of the data distribution without being heavily influenced by extreme values.

4. Mean Deviation

Mean deviation is a fundamental statistical measure that provides valuable insights into the dispersion. It serves as an essential tool for understanding the variability of a set of values, helping us to gauge how closely data points cluster around the mean of the dataset. In essence, mean deviation quantifies the average absolute difference between each data point and the mean, making it a versatile and informative statistic used in various fields, from finance and economics to science and engineering. In order to find the mean deviation, the following points have to be taken into account.

1. The mean deviation is generally used to find out the mean deviation.

2. If a term value (say **X**) is to find the deviation from the chosen numeric mean (say median **M**), then the value of the median in that term value is subtracted. That is, the deviation or drop of X value from the median to Deviation d = X-M.

3. Mean deviation is an absolute measure and its relative measure is called the mean-deviation factor.

Important methods to calculate mean deviation are:

I- Calculation of Mean Deviation in Individual Series

(a) Direct Method

Among the two methods of finding the mean deviation in an individual data series, the calculation of the direct method is completed in the following steps.

1 First of all, a statistical mean has to be selected and its value in the data series has to be found.

The value of mean deviation should be found by dividing the number of items in the sum (N) of the deviations obtained from 2 numerical series.

The main formulas used in statistical mean selected in personal data category are as follows.

1. Mean deviation based on arithmetic mean

2. Deviation based on median

Example: The area of a crop given below has been displayed in hectares; find its median and arithmetic mean, mean deviation and mean deviation coefficient by direct method.

: 4 8 6 7 15 18 21 23 26

Solution- First of all, the median and arithmetic mean will be determined by arranging the above data in ascending order and after this, the mean deviation and mean deviation coefficient from these statistical means will be calculated as per the table.

Mean-deviation coefficient calculation by direct method (Table No. 1.1)

Serial	Area in Hectares	Median i.e. 15 to ± Deviation	Deviation from arithmetic mean (i.e. 18) to ± Deviation
	X	**[X - M]or$\lvert d_M \rvert$**	**[X -$\bar{X}$]or$\lvert d\bar{X} \rvert$**
1	4	11	14
2	8	7	10
3	6	9	12
4	7	8	11
5	15	0	3
6	18	3	0
7	21	6	4
8	23	8	5
9	26	11	8
N= 9	$\sum \boldsymbol{X}$**= 128**	$\sum\lvert \boldsymbol{d_M} \rvert$ **= 63**	$\sum\lvert \boldsymbol{d\bar{X}} \rvert$**= 67**

¼1½ Calculation by median: Median = $\frac{\mathbf{9+1}}{\mathbf{2}}$ = **5** = term = **15** hectare

∴ Mean-deviation from the median

$$\delta M^{\cdot} = \frac{\sum|d_M|}{N} = \frac{63}{9} = \mathbf{7} \text{ Hectare}$$

And the mean-deviation coefficient $= \frac{\delta M}{M} = \frac{7}{15} = \mathbf{0.466}$

(2) Calculation according to the parallel mean:

Mean-deviation $\bar{X}\frac{\sum X}{N} = \frac{128}{9} = \mathbf{14.22}$ Hectare

∴ Mean-deviation from the deviation mean

$$\boldsymbol{\delta\,\overline{X}} = \frac{|d\overline{X}|}{N} = \frac{67}{9} = \mathbf{7.44} \text{ Hectare}$$

And the mean deviation coefficient from the deviation mean

$$\boldsymbol{\delta\,\overline{X}} = \frac{\delta\,\overline{X}}{\overline{X}} = \frac{7.44}{18} = \mathbf{0.41} \text{ Hectare}$$

(b) Short-cut Method

In individual data series, the mean deviation from the median or arithmetic mean is also determined by the short method. If the mean deviation is to be found on the basis of median then the following procedure is adopted.

1. First of all, by arranging the given data in ascending order, the value of median M is determined with the help of the median formula $(N+1)/2$

2- 2. The mean deviation from the median has to be calculated according to the following formula.

$\boldsymbol{\delta}_{\mathrm{M}} = \frac{\sum XA - \sum XB}{N}$ N = total number of posts.

On the contrary, in the short method of finding the mean deviation from the arithmetic mean, the following formula is used.

$$\delta\,\overline{X} = \frac{\sum X_{A-} \sum X_{B-\overline{X}}\,(N_{A-N_B})}{N}$$

The abbreviations of the above formula mean as follows.

$\delta\,\overline{X}$ = Mean deviation from arithmetic mean.

$\delta\overline{X}$= The value of arithmetic mean series.

$\sum \boldsymbol{X_A}$ = The sum of term values greater than arithmetic mean.

$\sum X_B$= Number of terms having values greater than the arithmetic mean.

N_A= Total number of posts in the category.

N_B= Number of posts having values less than arithmetic mean.

$\dot{N}$= Total number of posts in the category.

Solve the equations given in the above example by a short method.

Area of Hectares: 4 8 6 7 15 18 21 23 26

Arranging the data given in the question in ascending order first.......

4 6 7 8 15 18 21 23 26

(1) Calculation from the median:

Median = $\frac{\boldsymbol{N+1}}{\boldsymbol{2}}$ = 9+1 = 10 $\frac{\boldsymbol{10}}{\boldsymbol{2}} = 5$ term = **15**

$\sum X_{A=}$ $18+21+23+26 = 88$

$\sum \mathrm{X}_{\mathrm{B}}$ $=4+6+7+8=25$

∴ Median to mean deviation

$\delta\ \bar{X} = \frac{\sum \boldsymbol{X_{A-}} \sum \boldsymbol{X_B}}{\boldsymbol{N}}\ \frac{\boldsymbol{88-25}}{\boldsymbol{9}} = \frac{\boldsymbol{63}}{\boldsymbol{9}}$ = 7 Hectare

And the median from the mean deviation coefficient

$$\frac{\delta M}{M} = \frac{7}{15} = 0.466$$

¼2½ Calculations for the mean deviation

Mean deviation or $\bar{X} = \frac{\sum X}{X}\ \frac{128}{9} = 14.22\ Hectear$

$\sum \mathrm{X}_{\mathrm{A}} = 21 + 23 + 26 = 70$

$\sum X_B$ $=4+6+7+8+15 =40$

N = 9; N_A = 3 & N_B = 5

∴ Arithmetic Mean to Mean Deviation

$\delta\bar{X} = \sum\frac{X_A - \sum X_{B-\bar{X}} \, (N_{A-N_B}}{N}$

$= \frac{70-40-15(3-5)}{9} = \frac{60}{9}$

= 6.66 Hectare

And the coefficient of mean-deviation from arithmetic mean,

$= \frac{\delta\bar{X}}{X} = \frac{6.66}{18}$ = 0.37 Hectare

II- Calculation of Mean Deviation in Discrete Series

(a) Direct Method

Segmented series consists of direct and minor methods of finding the mean deviation from a statistical mean (arithmetic mean, median or mode) whose formulae is actually used in individual series.

The direct method is calculated as follows.

(1) In the given data series, the mean deviation is first determined by finding the statistical mean.

(2) The mark less deviation of each item-value is calculated by subtracting the value of the known arithmetic mean or median to X values.

(3) Values of mean deviation

By dividing the value of mean deviation by the respective mean, the relative measure of mean deviation is found.

As an example (Table No. 02), find the median coefficient of radiation and the mean coefficient of radiation in the following data series by direct method?

Table number 1.2

Area (in hectares)	4	8	12	16	20	24	28
Number of cultivation farm	4	5	6	8	7	4	2

To find the mean-deviation coefficient of the data in the above table no. 02, first the median and arithmetic mean are found, then the mean deviations are calculated from these means and finally the coefficients of deviation are calculated by dividing the respective mean in the mean-deviations. Calculations are done.

Calculation of median and arithmetic mean in a descries series (Table no 1.3)

Area in hectares	**Numbers of Cultivated farm / Frequency**	**Cumulated Frequency**	**Product of Area and Frequency**
X	**F**	**Cf**	**f×x**
4	4	4	16
8	5	9	40
12	6	15	72
16	8	23	128
20	7	30	140
24	4	34	96
28	2	36	56
	N= 36		$\boldsymbol{\sum fx = 548}$

Median = $\frac{N=1}{2} = \frac{36+1}{2} = \mathbf{18.5}$ **term** = 16 Hectare

Mean Deviation $\bar{X} = \frac{\sum FX}{N} = \frac{548}{36} = 15.22\ Hectares$

Calculation of mean deviation in a discrete series (Table No 1.4)

Area	Frequency	Median = 16 s		Paramillary mean 15.22 s	
		∓ Deviation	Total Deviation	∓ Deviation	Total Deviation
X	F	$\lceil d_M \rceil$	F$\times$ $\lvert d_M \rvert$	$d\lvert\bar{X}\rvert$	f$\times$ $\lvert d\bar{X}\rvert$
4	4	12	48	11.22	44.88
8	5	8	40	7.22	36.10
12	6	4	24	3.22	19.32
16	8	0	0	0.78	6.24
20	7	4	28	4.78	33.46
24	4	8	32	8.78	35.12
28	2	12	24	12.78	25.56
	N= 36	$\sum f\ \lvert d_M \rvert = 196$		$\sum f \lvert d\bar{X} \rvert$= 200.68	

Mean deviation from the median

$$\delta M = \frac{\sum f\ \lvert d_M \rvert}{N} = \frac{196}{36} = 5.444 \text{ Hectare}$$

$$\therefore Meadian\ coefficient\ of\ dispersion$$

$$= \delta \frac{M}{M} = \frac{5.444}{16} = 0.3402$$

∴ Mean deviation from arithmetic mean

$$= \delta\bar{X} = \sum f \frac{d\bar{x}}{N} = \frac{200.68}{36} = 5.574\ hectare$$

$$\therefore mean\ coefficient\ of\ dispersion \text{ c of } \delta\bar{X}$$

$$= \frac{\delta\bar{X}}{\bar{X}} = \frac{5.574}{15.22} = 0.3662$$

5. Standard Deviation

The square root of the arithmetic mean of the squares of deviations of various term values of a data series from the arithmetic mean is called the standard deviation of that series. Standard deviation is also called square-root deviation from arithmetic mean. Regarding the calculation of standard deviation, two things are kept in mind, (1) the mean deviation of a series can be calculated on the basis of any statistical mean.

(2) Like mean deviation, instead of ignoring the algebraic signs, every deviation extracted from the arithmetic mean is squared, as a result of which the value of the squares of negative deviations automatically becomes positive.

Coefficient of standard deviation $(C.of\ \sigma) = \frac{\sigma}{\bar{X}}$

(1) Calculation of Standard Deviation in Individual Series

There are two methods of finding the standard deviation in individual categories.

(1) Direct method (2) Indirect method

If the arithmetic mean of individual values is in integers then the direct method is used. On the contrary, if the value of the arithmetic mean is in decimal then the short method is used to simplify the calculation.

(2) Direct Methods

In the following individual data series **(Table No. 1.5),** calculate the standard deviation and standard deviation coefficient by direct and short method.

Observation Centre	Annual Rainfall (cm)
Pantnagar	34
Mukteswar	38
Hawalbag	40
Bageswar	43
Deharadun	45
Pauri Garhwal	46

Garun Ganga	48
Gopeswar	46

The standard deviation of the above data will be solved in the following way.

Direct calculation method of standard deviation in individual category (Table No. 1.6)

Observation Centre	**Annual Rainfall (cm)**	$\bar{X}$ $= \textbf{Deviation with } 42.5$	**Square of deviation**
	X	$(\times - \bar{X})or\ d$	(x- $\bar{X}$)^2 or d^2
Pantnagar	34	-8.5	72.25
Mukteswar	38	-4.5	20.25
Hawalbag	40	-2.5	6.25
Bageswar	43	+0.5	0.25
Deharadun	45	+2.5	6.25
Pauri Garhwal	46	+3.5	12.25
Garun Ganga	48	+5.5	30.25
Gopeswar	46	+3.5	12.25
N= 8	$\sum X = 340$		$\sum(X-\overline{X})^2 = \sum d^2 = 160$

Standard Deviation $\bar{X} = \frac{\sum X}{N} = \frac{340}{8} = 42.5\ cm.$

(1) ∴ Standard Deviation or $\sigma = \sqrt{\frac{\sum d^2}{N}} = \sqrt{\frac{160}{8}}$

$= \sqrt{20} = 4.47$ Cm

(2) Standard deviation coefficient or $(\text{C}\,.\,of\ \sigma) = \frac{\sigma}{\bar{X}}\frac{4.47}{42.5} = 0.105$

(B) Short-cut Method

Short method i.e. short methods of finding the standard deviation in an individual data series are also divided into two parts: 1. Value-group method and 2. Assumed mean method, but here we calculate the standard deviation of the value-group method.

1. First of all, we find the sum of individual values i.e. the value of $\sum X$.

2. After squaring each individual value, the value of $\sum X^2$ has to be found by adding these squares.

3. After that the value of standard deviation is calculated according to the following formula.

Standard deviation or $\sigma = \sqrt{\frac{\sum X^2}{N} - \left\{\frac{\sum X^2}{N}\right\}}$

Example: Find the standard deviation of the following data according to the value square method?

Calculation of Standard Deviation by Price Square Method **(Table No. 1.7)**

Observation Centre	**Annual Rainfall (cm.)**	**Square of price**
	X	**X^2**
Pantnagar	34	1156
Mukteswar	38	1444
Hawalbag	40	1600
Bageswar	43	1849
Deharadun	45	2025
Pauri Garhwal	46	2116
Garun Ganga	48	2304
Gopeswar	46	2116
N= 8	$\sum X = 340$	$\sum X^2 = 14610$

Standard deviation according to formula, $\sigma = \sqrt{\frac{\sum X^2}{N} - \left\{\frac{\sum X^2}{N}\right\}}$

$$\sqrt{\frac{14610}{8}} = \left\{\frac{340^2}{8}\right\}$$

$$\sqrt{1826.25-(42.5^2)}$$

$$\sqrt{1826.25-1806.25}$$

$$\sqrt{20} = 4.47\ cm.$$

II- Calculation of Standard Deviation in Discrete Series

Even in broken series, direct and short methods of finding standard deviation are used.

(a) Direct Method

In this method, the calculation process is done as follows (1) First of all, the arithmetic mean of the data series is found by dividing the total sum of values by the total number of terms.

(2) To find the deviation of each value from the arithmetic mean according to the ¼x-$\bar{X}$½ formula and to find the square (($\sum \boldsymbol{fX}$)) of these deviations.

(3) To find the value of standard deviation using the following formula.

Standard Deviation or $\sigma = \sqrt{\frac{\sum fd^2}{N}}$

Find the standard deviation and standard deviation coefficient in the following data series (Table No. 08) with the help of direct method.

Table No. 1.8

Wheat production in kg.	Number of families
8	4
9	5
10	9
11	20
12	18

13	8
14	6

On solving the above data by direct method… Table No. 1.9

Wheat production in Kg	Numbers of Family's/ Frequency	Production of output and frequency	Deviation from arithmetic mean 11.3	Square of deviation	Multiplication of square of deviation and frequency
X	F	f×x	$(\mathcal{X} \times \bar{X})$ or d	d^2	f×d^2
8	4	32	-3.3	10.89	43.56
9	5	45	-2.3	5.29	26.45
10	9	90	-1.3	1.69	15.21
11	20	200	-0.3	0.09	1.80
12	18	216	+0.7	0.49	8.82
13	8	104	+1.7	2.89	23.12
14	6	84	+2.7	7.29	43.74
	N=70	$\sum fX = 791$			$\sum fd^2 = 162.7$

Mean deviation $\bar{X}=\frac{\sum fX}{N} = \frac{791}{70} = 11.3$

$\therefore$ Standard deviation $\sigma = \sqrt{\frac{\sum fd^2}{N}}$

$= \sqrt{\frac{162.7}{70}} = \sqrt{2.324}$

= 1.524 m.ton

And standard deviation coefficient c. of $\sigma = \frac{\sigma}{\bar{X}} = \frac{1.524}{11.3} = 0.135$

(b) Short-cut Method

Short-cut method: Like individual method, short methods of finding standard deviation in broken series have been divided into two parts: 1. Short method based on classes of values and

2. Short method based on assumed mean. Here we will use a shortcut method based on squares of values.

Calculate the standard deviation in the following table no 09 divided series data based on value Square method? In short-cut method, to calculate the value by square method, first of all the following procedure has to be adopted.

1) Multiply each value (x) by its frequency (f) and add the obtained products to get the value of (∑fx).

2) find the total number of terms ¼∑f or n) by adding all the frequencies.

3) Find the value of ($\sum x^2$) by multiplying each value's frequency (f) in the square of ($\sum x^2 f$½.

4) If the standard deviation coefficient is also to be found, then the arithmetic mean (¼∑fx/n½ can be found through the formula ¼$\bar{X}$½.

Finally the standard deviation is calculated using the following formula.

$$\text{Standard deviation or } \sigma = \sqrt{\frac{\sum X^2 f}{N} - \left\{\frac{\sum f X}{N}\right\}^2}$$

$$\text{or } \sigma = \sqrt{\frac{\sum X^2 f}{N} - \{\bar{X}\}^2}$$

Solve

Table No- 1.10

Wheat production in Kg.	Numbers of Family's/ Frequency	Production of output and frequency	Square of Production	Multiplication of square of production and frequency
x	**F**	**f×x**	**x^2**	**x^2×f**
8	4	32	64	256
9	5	45	81	405
10	9	90	100	900
11	20	200	121	2420
12	18	216	144	2592

13	8	104	169	1352
14	6	84	196	1176
	N=70	$\sum$fX= 791		$\sum X^2 f$=9101

Mean deviation or $\bar{X} = \frac{\sum fX}{N} = \frac{79}{70}$ = 11.3

Standard deviation $\sigma \sqrt{\frac{\sum X^2 f}{N} - \bar{X}^{\ 2}}$

= $\sqrt{\frac{9101}{70} - (11.30)^2}$

=$\sqrt{130.014 - 127.69}$ = $\sqrt{2.324}$

= 1.524 m. ton

And standard deviation coefficient c. of $\sigma = \frac{\sigma}{\bar{X}} = \frac{1.524}{11.3}$ = 0.135

(ii) Assumed Mean Method

This method is based on the deviations extracted from the assumed mean and in this

(1) The standard deviation is calculated from the assumed mean (**A**) of the given values.

(2) The deviation of each value from this assumed mean **X-A** or dx the jug is taken out.

(3) The square or $\mathbf{dx^2}$ of each deviation also has to be calculated.

(4) In this way, the value of $\sum \mathbf{fdx^2}$ is found by adding the products obtained by multiplying the respective frequencies in the squares of deviations.

(5) After that the value of standard deviation is estimated.

Table number 10 explains the calculation of standard deviation in a divided series.

Table Number (1.11)

Wheat production in Kg.	**Numbers of Family's/ Frequency**	**Production of output and frequency**	**Deviation (A=10) from the Assumed Mean**	**Square of deviation**	**Multiplication of square of production and frequency**	**Product of frequency and square of deviation**
X	**f**	**Fx**	**(X-A)**	**dx^2**	**Fdx**	**fdx^2**
8	4	32	-2	4	-8	16
9	5	45	-1	1	-5	5
10	9	90	0	0	0	0
11	20	200	+1	1	+20	20
12	18	216	+2	4	+36	72
13	8	104	+3	9	+24	72
14	6	84	+4	16	+24	96
	N=70	$\sum fX = 791$			$\sum fdx = 91$	$\sum fdx^2 = 281$

Standard deviation $\sigma = \sqrt{\frac{\sum f d^2 x}{N} - \left\{\frac{\sum f dx}{N}\right\}^2}$

$= \sqrt{\frac{281}{70} - \left\{\frac{91}{70}\right\}^2}$

$= \sqrt{4.014 - (1.3)^2}$

$= \sqrt{4.014 - 1.69} = \sqrt{2.324}$

= 1.524 m. ton

V- Utility of Standard Deviation

Standard deviation is a definite, clear and ideal measure of dispersion which is more useful than the dispersion of two or more data series. In the measurements of central tendency, the same place is given to the arithmetic mean; the same place is given to the standard deviation in the measurements of dispersion.

1.4 SUMMARY

In statistical science, statistical methods related to collection, classification, presentation, analysis, comparison and interpretation of data are mainly studied. Whose objective is to reach an appropriate conclusion by analyzing numerical facts and this is possible only when the methods used in analyzing numerical facts etc. are selected judiciously. Drawing conclusions through statistical methods has become a modern trend in all subjects. In the last 2-3 decades, the basic nature of geography has changed and statistics has started being used in the problem of basic study material of the subject. Which has started displaying the quantitative characteristics of the subject more easily instead of the qualitative characteristics and after the quantitative revolution, it is studying the geographical elements and events with more precision and accuracy in relation to the country and time. Thus, the usefulness and importance of statistical data in the subject of geography is increasing day by day, which is becoming more successful in clarifying complex descriptive facts through mathematical methods.

1.5 GLOSSARY

- **Statistical data:** Those data of statistical science which do not represent just a single number but represent the totality of a group.
- **Primary data:** Statistical data that a user collects himself. Such as interview, survey data.
- **Secondary data:** are those data which have been compiled by any institution or government department, such as census.
- **Dispersion:** average difference of different individual values of a data series from the mean of the series is called dispersion.
- **Range:** The difference between the largest value and the smallest value in a data series is called the extension or range of that series.
- **Mean Deviation:** The arithmetic mean or mode or arithmetic mean of the deviations of different term values from the median series is known as mean deviation.
- **Standard Deviation:** The square root of the arithmetic mean of the squares of deviations of its various term values from the arithmetic mean of a data series is called the standard deviation or standard deviation of that series.

1.6 ANSWER TO CHECK YOUR PROGRESS

1. Range the difference between the largest and smallest term values in a data series is called range? Which is called R=L-S

2. Radiation coefficient series reveals the average difference of various individual values from the mean of the series?

3. Limit method, deviation mean method and graphical method are included in the measurement of radiation?

4. The coefficient of expansion is determined by (C or R) = $\frac{L-S}{L+S}$?

5. What is the simplest method of measuring infrared radiation which is most important in displaying the data of annual rainfall i.e. rainfall of a place?

6. To find the arithmetic mean, first of all, it is determined by the deviation of the term values of any statistical mean, arithmetic mean, mode or median of the data series?

7. σ is a word of the Greek alphabet which is also called delta?

8. The relative measure of mean deviation is called mean deviation coefficient?

9. Standard deviation is also called the square-mean-root deviation from the arithmetic mean?

10. Standard deviation is a fixed, ideal measure of radiation?

1.7 REFERENCES

- Sharma, J.P. : 2007-2008, Practical Geography, Rastogi Publications Gangotri Shivaji Road, Meerut.
- Mishra, P.L. : 2010 Practical Geography, Vishavabharti Publication New Delhi.

1.8 TERMINAL QUESTIONS

(i) Long Type Question

Q.1-What is mean and standard deviation? With the help of the following data, calculate the mean deviation by direct and short method?

8, 7, 6, 11, 10, 16, 15 & 17

Q.2- Calculate the standard deviation coefficient in the following individual data series by direct method?

Table No 1.12

Observation Center	Annual Rainfall (cm) X
Pantnagar	34
Mukteswar	38
Hawalbag	40
Bageswar	43
Deharadun	45
Pauri Garhwal	46
Garun Ganga	48
Gopeswar	46

(ii) Short Type Questions

Question 1: What do you understand by measurement of radiation?

Question 2: What is extension or range?

Question 3: Write the formula for calculating range?

Question 4: Explain the utility of range?

Question 5 what is meant by mean deviation?

Question 6: Write the formulas of mean deviation in individual series?

Question 7: Explain the utility of mean deviation?

Question 8: What do you understand by standard deviation?

Question 9: Explain the formulas of standard deviation in individual category?

Question 10: Explain the short method based on assumed mean?

UNIT- 2 SAMPLING THEORY: BASIC CONCEPTS OF PROBABILITY, TEST & SIGNIFICANCE OF 'T' TEST 'CHI-SQUARE' TEST, CENTRAL TENDENCY MEAN, MEDIAN, MODE, CENTRAL TENDENCY

2.1 OBJECTIVES

2.2 INTRODUCTION

2.3 MEASURE CENTRAL TENDENCY MEAN, MEDIAN, MODE

2.4 BASIC CONCEPTS OF PROBABILITY

2.5 TEST & SIGNIFICANCE'T' TEST & 'CHI-SQUARE' TEST

2.6 SUMMARY

2.7 GLOSSARY

2.8 ANSWER TO CHECK YOUR PROGRESS

2.9 REFERENCES

2.10 TERMINAL QUESTIONS

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the sampling theory.
- Learn about the Basic concepts of probability.
- Gain knowledge about test & Significance ‘t’ test ‘chi-square’ test.
- Learn about the central tendency Mean, Median, Mode, and Central tendency.

2.2 INTRODUCTION

A sample is a small part of data taken from a larger group that's available. Sampling is important because it's hard to gather data from every single source. For instance, if you want to interview everyone in a town, that's usually not possible. Instead, you choose a group of people to represent the town. The way you pick this group is called the sampling technique.

Sampling not only makes collecting data possible but also saves time. Usually, it's not needed to get data from every single source, and sometimes we don't even know the total population size or where exactly to collect data from. This uncertainty can make it hard for researchers to know when to stop collecting data.

Researchers need to put in a lot of effort to pick the right sample size for their data. Bigger sample sizes usually reflect the sampling frame better, making the results more reliable. However, researchers also need to consider how manageable it is to collect and handle the data, and the time and resources available, when deciding on the sample size.

There are different ways to sample, but three popular methods are random, systematic, and stratified sampling. Let's see how these work with two examples.

First, imagine Researcher A has to pick ten spots for data collection from a map divided into three zones. Then, there's Researcher B who needs to choose ten interviewees from a group of twenty people of different ages.

Purpose of Sampling

Sampling is crucial in geography for collecting data and drawing informed conclusions about geographic phenomena. It means picking some locations, regions, or areas from a big geographic area to stand for the whole thing we're interested in. This is important because it lets geographers study lots of different geographic things, like patterns of land cover and how people settle in different places, without having to look at every single bit of the Earth's surface.

The main goal of sampling in geography is to get a group of locations or areas that shows what the bigger geographic area is like accurately. By choosing these places carefully, researchers try to include all the different environmental, social, and economic things that make up the landscape. This group of places is like a small version of the whole area, so geographers can study how things are spread out, find trends, and understand what's happening in the whole area.

Types of Sampling

1. **Subjective sampling**
2. **Objective sampling**

Subjective sampling

Subjective sampling, also called judgmental or purposive sampling, is a way researchers gather data by using their judgment and knowledge to pick specific participants, places, or samples based on set criteria. Unlike random sampling, where everyone has an equal chance of being chosen, subjective sampling is deliberate and selective, focusing on subjects that best fit the research goals.

One big thing about subjective sampling is that the researcher's judgment guides the selection process. Instead of relying on chance or math to make sure the sample represents the whole group, the researcher uses their expertise to choose individuals or cases they think will give the most useful insights. This method is often used in qualitative research, where the aim is to understand complex social stuff and explore lots of different views.

Types of Subjective Sampling

1. **Accidental Sampling**
2. **Quota Sampling**
3. **Purposive Sampling**
4. **Convenience Sampling**

Accidental Sampling

Quota Sampling

In quota sampling, people are chosen based on certain standards or criteria set beforehand. This method makes sure the sample has certain traits that are just like those in the whole population. So, the sample looks like the whole population in terms of those traits. Quota sampling is a fast and effective way to gather samples because it lets researchers collect data quickly while still making sure the sample represents important traits of the population.

Purposive Sampling

Judgmental or purposive sampling relies on the researcher's judgment. They choose participants who align with the study's purpose and their grasp of the intended audience. For instance, if researchers want insights into the thinking of people considering a master's degree, they might use a question like, "Are you interested in pursuing a master's in ...?" Those who answer "No" wouldn't be included. This way, the sample includes individuals who are most pertinent to the research goals.

Convenience Sampling

This method relies on how easy it is to reach subjects, like surveying shoppers at a mall or people passing by on a busy street. It's called convenience sampling because it's simple for researchers to do and contact subjects. Researchers don't have much control over who gets picked because it's based on who's nearby, not if they represent the whole group. This type of sampling, which doesn't involve probability, is used when time and money are tight. Convenience sampling is often used when resources are limited, like in the early stages of research. For instance, startups and NGOs might use it at a mall to hand out leaflets for events or to spread awareness. They'd just stand at the entrance and give them out randomly.

Objective Sampling

Types of Objective Sampling

1. **Random Sampling**
2. **Stratified Random Sampling**
3. **Systematic Sampling**
4. **Cluster Sampling**

Random Sampling

Random sampling involves picking data sources in a completely unplanned way. Once the sample size is set, usually as a percentage of the total group, researchers use random number generators, which you can find online easily, to make totally random sets of numbers. These numbers can then help pick spots on a map for data collection or identify specific houses to survey on a street.

If a researcher needs a random line on a map, a random number generator can give the starting and ending points for that line. For surveys in nature, where they're using a quadrat, a common way is to stand in the middle and throw the quadrat with eyes shut, then survey where it lands. Doing this again from the new spot makes a random set of spots.

Other ways to make random numbers include rolling dice, picking playing cards without looking, or choosing bingo numbers from a bag. These methods take away human choice from the process, which helps lower the chance of getting biased results. But random sampling might not work well for small groups because there aren't many choices available.

Stratified Random Sampling

Stratified sampling splits the sampling frame into smaller groups, called strata, and then uses these groups to adjust the sample to reflect the original frame accurately. For instance, if 30% of the frame is from a specific location, age group, or religion, then 30% of the sample will also represent these groups.

Many researchers see stratified sampling as the fairest method because it makes sure the sample mirrors the original frame well. But it needs good knowledge of the frame to pick the right groups, and sometimes a trial study is needed to figure out how much each group should count.

Depending on the research, different sampling methods can be mixed. For example, a researcher might draw a random line across a data collection site on a map and then use systematic or stratified sampling to pick exact collection points.

Systematic Sampling

Systematic sampling picks data sources in a structured, but not random, way. The sample size might not be set beforehand because the sampling system itself decides it. The researcher picks a gap size between samples—like a set distance on a map or picking every nth person in a survey—and sticks to it without changing.

Systematic sampling is good because it removes the researcher from the selection process, which lowers the chance of bias. But to make sure the sample is like the study area, the researcher needs to be careful that the sampling frame doesn't add bias. For example, if the frame comes from a list of people, like the electoral roll, it would leave out people who can't vote, like those under eighteen and some prisoners.

Cluster Sampling

Cluster sampling involves splitting the whole population into groups or clusters that represent it. These clusters are chosen for the sample based on demographic factors like age, gender, location, and so on. This method makes it easier for survey makers to get useful insights from the feedback.

For instance, if the US government wants to know how many immigrants are in the mainland, they could split the population into clusters based on states like California, Texas, Florida, Massachusetts, Colorado, Hawaii, and so on. Doing the survey like this would give organized results by state, giving useful immigration info.

2.3 MEASURE CENTRAL TENDENCY MEAN, MEDIAN, MODE

In statistics, measures of central tendency are single values aiming to describe a dataset by pinpointing its central position. These measures offer insights into where the "centre" of the data lies. Three commonly used measures of central tendency include the mean, median, and mode.

Mean

The mean, also known as the average, is calculated by summing all the values in the dataset and then dividing by the total number of values.

Example: Consider the following dataset of exam scores: {60, 70, 80, 85, 90}. The mean ($\bar{x}$) is calculated as:

$$\bar{x} = \frac{60 + 70 + 80 + 85 + 90}{5} = 77$$

Arithmetic Mean

The arithmetic mean, often simply called the mean, is the sum of a collection of numbers divided by the count of numbers in the collection. It is the most commonly used measure of central tendency.

Definition

For a set of numbers $x_1, x, \dots, x$, the arithmetic mean $\bar{x}$ is given by:

$$\bar{x} = \frac{1}{x}\sum_{i=1}^{n} x_i$$

Example

Consider the set of numbers 2,3,5,7,11. The arithmetic mean is calculated as follows:

$$\bar{x} = \frac{1}{5}(2 + 3 + 5 + 7 + 11) = \frac{1}{5} \cdot 28 = 5.6$$

Arithmetic Mean with Frequency

In some cases, each number in the data set may occur with a certain frequency. The arithmetic mean can be calculated by taking into account these frequencies.

Example

Consider a set of test scores 50,60,70,80,90 with frequencies 2,3,5,2,1 respectively. The arithmetic mean can be calculated as follows:

$$\mathrm{x} = \frac{(50 \times 2) + (60 \times 3) + (70 \times 5) + (80 \times 2) + (90 \times 1)}{2 + 3 + 5 + 2 + 1} = \frac{100 + 180 + 350 + 160 + 90}{13}$$

$$= \frac{880}{13} \approx 67.69$$

Table No. 2.1 Cricketers' Scores

Cricketer	Innings 1	Innings 2	Innings 3	Innings 4	Innings 5
S. Dhawan	45	60	30	70	50
R. Sharma	90	30	60	80	40
A. Rahane	50	70	20	50	30
V. Kohli	100	90	60	70	80
M. S. Dhoni	30	40	50	60	50
R. Jadeja	40	30	20	10	50
D. Karthik	60	50	40	30	20

Calculations of Arithmetic Mean

Individual Means

$$\mathrm{x} = \frac{45 + 60 + 30 + 70 + 50}{5} = 51$$

$$\bar{x}_{\text{R. Sharma}} = \frac{90 + 30 + 60 + 80 + 40}{5} = 60$$

$$\bar{x}_{\text{A. Rahane}} = \frac{50 + 70 + 20 + 50 + 30}{5} = 44$$

$$\bar{x}_{\text{V. Kohli}} = \frac{100 + 90 + 60 + 70 + 80}{5} = 80$$

$$\bar{x}_{\text{M. S. Dhoni}} = \frac{30 + 40 + 50 + 60 + 50}{5} = 46$$

$$\bar{x}_{\text{R. Jadeja}} = \frac{40 + 30 + 20 + 10 + 50}{5} = 30$$

$$\bar{x}_{\text{D. Karthik}} = \frac{60 + 50 + 40 + 30 + 20}{5} = 40$$

Overall Mean

The overall mean score across all innings and all cricketers can be calculated as:

$$x_{overall} = \frac{45+60+30+70+50+90+30+60+80+40}{35} + \frac{50+70+20+50+30+100+90+60+70+80}{35} + \frac{30+40+50+60+50+40+30+20+10+50+60+50+40+30+20}{35}$$

$$= \frac{1485}{35} = 42.43$$

Table No. 2.2 Cricketers' Scores with Frequencies

Cricketer	Score	Friquency
S. Dhawan	45,60, 30, 70, 50,	3, 1, 4, 2, 1,
R. Sharma	90, 30, 60, 80, 40,	2, 5, 1, 0, 4,
A Rahane	50, 70, 20, 50, 30,	3, 2, 1, 0, 5,
V Kohali	100, 90, 60, 70, 80,	1, 2, 0, 3, 2,
M. S. Dhoni	30, 40, 50, 60, 50,	4, 1, 5, 0, 3,
R Jadeja	40, 30, 20, 10, 50,	2, 3, 1, 5, 0,
D Karthik	60, 50, 40, 30, 20,	4, 2, 0, 1, 3,

Calculation of Arithmetic Mean for Each Cricketer

S. Dhawan

$$\bar{x}_{\text{Dhawan}} = \frac{(45 \times 3) + (60 \times 1) + (30 \times 4) + (70 \times 2) + (50 \times 1)}{3 + 1 + 4 + 2 + 1}$$
$$= \frac{135 + 60 + 120 + 140 + 50}{11}$$
$$= \frac{505}{11}$$
$$= 45.91$$

R. Sharma

$$\bar{x}_{\text{Sharma}} = \frac{(90 \times 2) + (30 \times 5) + (60 \times 1) + (80 \times 0) + (40 \times 4)}{2 + 5 + 1 + 0 + 4}$$
$$= \frac{180 + 150 + 60 + 0 + 160}{12}$$
$$= \frac{550}{12}$$
$$= 45.83$$

A. Rahane

$$\bar{x}_{\text{Rahane}} = \frac{(50 \times 3) + (70 \times 2) + (20 \times 1) + (50 \times 0) + (30 \times 5)}{3 + 2 + 1 + 0 + 5}$$
$$= \frac{150 + 140 + 20 + 0 + 150}{11}$$
$$= \frac{460}{11}$$
$$= 41.82$$

V. Kohli

$$\bar{x}_{\text{Kohli}} = \frac{(100 \times 1) + (90 \times 2) + (60 \times 0) + (70 \times 3) + (80 \times 2)}{1 + 2 + 0 + 3 + 2}$$
$$= \frac{100 + 180 + 0 + 210 + 160}{8}$$
$$= \frac{650}{8}$$
$$= 81.25$$

M. S. Dhoni

$$\bar{x}_{\text{Dhoni}} = \frac{(30 \times 4) + (40 \times 1) + (50 \times 5) + (60 \times 0) + (50 \times 3)}{4 + 1 + 5 + 0 + 3}$$
$$= \frac{120 + 40 + 250 + 0 + 150}{13}$$
$$= \frac{560}{13}$$
$$= 43.08$$

R. Jadeja

$$\bar{x}_{\text{Jadeja}} = \frac{(40 \times 2) + (30 \times 3) + (20 \times 1) + (10 \times 5) + (50 \times 0)}{2 + 3 + 1 + 5 + 0}$$
$$= \frac{80 + 90 + 20 + 50 + 0}{11}$$
$$= \frac{240}{11}$$
$$= 21.82$$

D. Karthik

$$\bar{x}_{\text{Karthik}} = \frac{(60 \times 4) + (50 \times 2) + (40 \times 0) + (30 \times 1) + (20 \times 3)}{4 + 2 + 0 + 1 + 3}$$
$$= \frac{240 + 100 + 0 + 30 + 60}{10}$$
$$= \frac{430}{10}$$
$$= 43.00$$

Geometric Mean

The geometric mean is more appropriate for sets of numbers whose values are meant to be multiplied together or are exponential, such as rates of growth.

Example

Consider the set of numbers 1,3,9,27,81. The geometric mean is calculated as follows:

$$G = (1 \times 3 \times 9 \times 27 \times 81)^{\frac{1}{5}} = 3^2 = 9$$

Harmonic Mean

The harmonic mean is useful for sets of numbers that are defined concerning some unit, such as speeds or rates.

Example

Consider the set of numbers 1,2,4. The harmonic mean is calculated as follows:

$$\mathrm{H} = \frac{3}{\frac{1}{1}+\frac{1}{2}+\frac{1}{4}} = \frac{3}{1+0.5+0.25} = \frac{3}{1.75} \approx 1.714$$

Median

The median is the middle value of a dataset when it is ordered from least to greatest. If there is an even number of data points, the median is the average of the two middle values.

Example: Using the same dataset of exam scores: {60, 70, 80, 85, 90}. The median is 80, as it is the middle value when the scores are ordered.

Mode

The mode is the value that appears most frequently in a dataset. A dataset can have one mode, more than one mode (multimodal), or no mode (no value appears more than once).

Example: Using the same dataset of exam scores: {60, 70, 80, 85, 90}. There is no mode as all values appear only once.

The concept of the “mean” is fundamental in mathematics and statistics. It represents a central value or typical value for a set of numbers. There are several types of means, each with different applications and interpretations. The most common types are the arithmetic mean, geometric mean, and harmonic mean. Here, we will describe each type and provide example

Conclusion

Each type of mean offers a distinct approach to summarizing a set of numbers, and the selection of the appropriate mean hinges on the data's characteristics and the context of its application. The arithmetic mean is well-suited for general purposes, the geometric mean for scenarios involving multiplicative processes, and the harmonic mean for rates and ratios.

Measure of Dispersion

In statistics, a measure of dispersion is a numerical value that describes how spread out or dispersed the values in a dataset are. It provides information about the variability or diversity of the data points.

There are several common measures of dispersion, including the range, variance, and standard deviation.

Range

The range is the simplest measure of dispersion. It is calculated as the difference between the maximum and minimum values in a dataset.

Example: Consider the following dataset of exam scores: $\{60, 70, 80, 85, 90\}$. The range is calculated as $90 - 60 = 30$.

Variance

Variance measures the average squared deviation of each data point from the mean of the dataset.

Example: Using the same dataset of exam scores: {60, 70, 80, 85, 90}, the mean ($\bar{x}$)is calculated as $(60 + 70 + 80 + 85 + 90)/5 = 77$. The variance ($s^2$) is calculated as:

$$s^2 = \frac{(60-77)^2 + (70-77)^2 + (80-77)^2 + (85-77)^2 + (90-77)^2}{5}$$

Standard Deviation

The standard deviation is the square root of the variance. It measures the average deviation of each data point from the mean.

Example: Using the same dataset of exam scores: {60, 70, 80, 85, 90}, the standard deviation (s) is calculated as the square root of the variance.

$$s = \sqrt{\frac{(60-77)^2 + (70-77)^2 + (80-77)^2 + (85-77)^2 + (90-77)^2}{5}}$$

2.4 BASIC CONCEPTS OF PROBABILITY

"Probability theory, a significant field within mathematics, focuses on the quantification of uncertainty and randomness. It offers a structured approach for modelling and scrutinizing uncertain events and their potential outcomes. Below are several fundamental concepts in probability theory:"

1. **Sample Space**: The sample space, denoted by S, is the set of all possible outcomes of a random experiment.
2. **Event**: An event is any subset of the sample space. It represents a collection of possible outcomes of interest.
3. **Probability**: The probability of an event A, denoted by $P(A)$, is a measure of the likelihood that A will occur. It is a number between 0 and 1, where 0 indicates impossibility and 1 indicates certainty.
4. **Probability Distribution**: A probability distribution describes the likelihood of each possible outcome of a random variable. It can be discrete or continuous.
5. **Random Variable**: A random variable is a variable whose possible values are outcomes of a random phenomenon. It can be discrete or continuous.

6. **Expected Value**: The expected value of a random variable X, denoted by $E(X)$ or μ, is the weighted average of all possible values of X, where the weights are the probabilities of the corresponding outcomes.

7. **Variance and Standard Deviation**: Variance measures the dispersion of a random variable's values around its mean, while standard deviation measures the average deviation of values from the mean.

8. **Independence**: Two events A and B are independent if the occurrence of one does not affect the probability of the other.

9. **Conditional Probability**: Conditional probability is the probability of an event A given that another event B has occurred, denoted by $P(A|B)$.

10. **Bayes' Theorem**: Bayes' theorem relates the conditional and marginal probabilities of two events. It is commonly used in statistical inference and machine learning.

Probability theory has applications in various fields, including statistics, economics, finance, computer science, and engineering.

Some Probability Examples

Example 1: Coin Toss

Let's consider the probability of getting heads when flipping a fair coin.

$$P(\text{Heads}) = \frac{\text{Number of favourable outcomes}}{\text{Total number of outcomes}}$$

Since there are 2 possible outcomes (heads or tails) and the coin is fair, the probability of getting heads is:

$$P(\text{Heads}) = \frac{1}{2} = 0.5$$

Example 2: Rolling a Die

What is the probability of rolling an even number with a fair six-sided die?

$$P(\text{Even number}) = \frac{\text{Number of favourable outcomes}}{\text{Total number of outcomes}}$$

There are 3 even numbers (2, 4, 6) on the die, and a total of 6 possible outcomes, so the probability of rolling an even number is:

$$P(\text{Even number}) = \frac{3}{6} = 0.5$$

Example 3: Drawing a Card

Suppose you draw a card from a standard deck of 52 playing cards. What is the probability of drawing a heart?

$$P(\text{Heart}) = \frac{\text{Number of hearts}}{\text{Total number of cards}}$$

There are 13 hearts in a standard deck of 52 cards, so the probability of drawing a heart is:

$$P(\text{Heart}) = \frac{13}{52} = 0.25$$

Example 4: Flipping a Coin Multiple Times

If you flip a fair coin three times, what is the probability of getting exactly two heads?

To calculate this probability, we use the binomial probability formula:

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where n is the number of trials, k is the number of successes, and p is the probability of success on each trial.

For this example, $n = 3$, $k = 2$, and $p = 0.5$ (since the coin is fair).

$$P(2\text{ heads}) = \binom{3}{2} (0.5)^2 (1-0.5)^{3-2} = 3 \times 0.25 \times 0.5 = 0.375$$

Example 5: Drawing Marbles from a Bag

Suppose you have a bag containing 5 red marbles and 3 blue marbles. If you draw two marbles without replacement, what is the probability of getting one red and one blue marble?

First, let's calculate the probability of drawing a red marble and then a blue marble:

$$P(\text{Red, then Blue}) = P(\text{Red}) \times P(\text{Blue after Red})$$

$$= \frac{5}{8} \times \frac{3}{7}$$

$$= \frac{15}{56}$$

Similarly, the probability of drawing a blue marble and then a red marble is also $\frac{15}{56}$.

Since these are mutually exclusive events, we can add their probabilities to get the total probability of getting one red and one blue marble:

$$P(\text{One Red and One Blue}) = 2 \times \frac{15}{56} = \frac{15}{28}$$

2.5 TEST & SIGNIFICANCE'T' TEST & 'CHI-SQUARE' TEST

Statistical tests serve as methodologies for concluding populations through the analysis of data samples. They aid researchers in assessing whether the evidence supports rejecting a null hypothesis in favour of an alternative hypothesis. An array of statistical tests exists, each tailored to specific data types and research inquiries. Below are several commonly employed ones:

t-Test

The t-test is used to determine if there is a significant difference between the means of the two groups. It is often used when the sample size is small and the population standard deviation is unknown.

Example:

Consider a study comparing the effectiveness of two teaching methods on exam scores. The t-test can be used to determine if there is a significant difference in the mean exam scores of students taught using Method A versus Method B.

Chi-Square Test

The chi-square test is used to determine if there is a significant association between two categorical variables. It compares the observed frequencies of the categories with the expected frequencies under the null hypothesis of independence.

Example:

A survey is conducted to investigate the relationship between gender and voting preference. The chi-square test can be used to determine if there is a significant association between gender (male/female) and voting preference (candidate A/candidate B).

ANOVA (Analysis of Variance)

ANOVA is used to determine if there is a significant difference in the means of three or more groups. It compares the variability between groups to the variability within groups to assess whether the group means are significantly different.

Example:

A study examines the effect of different doses of a drug on pain relief. ANOVA can be used to determine if there is a significant difference in pain relief scores among patients receiving different doses of the drug.

Linear Regression

Linear regression is used to model the relationship between a dependent variable and one or more independent variables. It estimates the parameters of a linear equation that best fits the data.

Example:

A researcher wants to predict students' exam scores based on the number of hours they study per week. Linear regression can be used to quantify the relationship between study hours (independent variable) and exam scores (dependent variable).

T-test

The t-test is a parametric statistical test used to determine if there is a significant difference between the means of two groups. It is commonly used when the sample size is small and the population standard deviation is unknown. There are several types of t-tests, including:

- Independent samples t-test: Compares the means of two independent groups.
- Paired samples t-test: Compares the means of two related groups, such as before and after treatment measurements.
- One-sample t-test: Tests whether the mean of a single sample differs from a known value.

The t-test produces a t-statistic and a p-value. The t-statistic measures the difference between the sample means relative to the variation in the data, while the p-value indicates the probability of observing the data if the null hypothesis (no difference between means) is true. A small p-value (typically less than 0.05) indicates evidence against the null hypothesis, suggesting a significant difference between the means.

Chi-square Test

The chi-square test is a non-parametric statistical test used to determine if there is a significant association between two categorical variables. It is often used to analyze contingency tables, where each cell represents the frequency of observations for combinations of categories of two variables. There are two main types of chi-square tests:

- Chi-square test for independence: Tests whether there is a relationship between two categorical variables. It compares the observed frequencies in the contingency table with the frequencies that would be expected if the variables were independent.
- Chi-square test for goodness of fit: Tests whether the observed frequency distribution of a single categorical variable fits an expected distribution. It is used when comparing observed frequencies to expected frequencies specified by a theoretical model.

Like the t-test, the chi-square test produces a chi-square statistic and a p-value. A small p-value (typically less than 0.05) indicates evidence against the null hypothesis, suggesting a significant association between the variables.

In summary, t-tests are used to compare means of continuous variables between two groups, while chi-square tests are used to assess the association between categorical variables.

t-Test

The t-test is a statistical technique utilized to ascertain whether a notable distinction exists between the averages of two groups. It finds frequent application in scenarios involving limited sample sizes where comparisons between the means of two populations are necessary.

Here's an example scenario:

Let's say you're a researcher studying the effect of a new drug on blood pressure. You have a group of 20 patients with high blood pressure, and you want to know if the new drug lowers their blood pressure significantly.

You divide the patients into two groups:

- Group A: 10 patients who receive the new drug.
- Group B: 10 patients who receive a placebo.

After a month of treatment, you measure the blood pressure of each patient and record the results.

Here's how you can perform a t-test to analyze the data:

11. Formulate Hypotheses:
    - Null Hypothesis (H0): There is no significant difference in the mean blood pressure between the two groups.
    - Alternative Hypothesis (H1): There is a significant difference in the mean blood pressure between the two groups.
12. Collect Data: Let's say the average blood pressure reduction in Group A is 10 mmHg, and in Group B, it's 2 mmHg.
13. Calculate Sample Statistics: Calculate the mean ($\bar{x}$) and standard deviation (s) of blood pressure in each group.
14. Determine the Significance Level (α): Typically set to 0.05, indicating a 5% chance of rejecting the null hypothesis when it's true.
15. Perform the t-test: Using the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{x}_1$ and $\bar{x}_2$ are the means of Group A and Group B respectively.

- s_1 and s_2 are the standard deviations of Group A and Group B respectively.
- n_1 and n_2 are the sample sizes of Group A and Group B respectively.

16. Interpret the Results: If the calculated t-value is greater than the critical value from the t-distribution table or if the p-value is less than α, then you reject the null hypothesis and conclude that there is a significant difference in blood pressure between the two groups.

t-Test Example

Suppose we have two sets of data representing the exam scores of two groups of students: Group A and Group B. We want to determine if there is a significant difference in the mean exam scores between the two groups.

Table no. 2.3 Data

Group A	Group B
78	84
82	79
85	80
80	83
75	86

Hypotheses

Let μ_A be the mean exam score of Group A and μ_B be the mean exam score of Group B.

Null Hypothesis (H_0): There is no significant difference between the mean exam scores of Group A and Group B.

$$H_0: \mu_A = \mu_B$$

Alternative Hypothesis (H_1): There is a significant difference between the mean exam scores of Group A and Group B.

$$H_1: \mu_A \neq \mu_B$$

Calculate Mean and Standard Deviation

Let's calculate the sample mean ($\bar{x}$) and sample standard deviation (s) for both groups.

Group A

$$\bar{x}_A = \frac{78 + 82 + 85 + 80 + 75}{5} = 80$$

$$s_A = \sqrt{\frac{(78-80)^2 + (82-80)^2 + (85-80)^2 + (80-80)^2 + (75-80)^2}{5-1}} \approx 3.41$$

Group B

$$\bar{x}_B = \frac{84 + 79 + 80 + 83 + 86}{5} = 82.4$$

$$s_B = \sqrt{\frac{(84 - 82.4)^2 + (79 - 82.4)^2 + (80 - 82.4)^2 + (83 - 82.4)^2 + (86 - 82.4)^2}{5 - 1}} \approx 2.88$$

2.6 SUMMARY

Sampling theory involves selecting a subset (sample) from a larger population to infer characteristics about the entire population. Central to this process are the basic concepts of probability, which provide the framework for understanding and calculating the likelihood of different outcomes. This theoretical foundation supports various statistical tests used to analyze sampled data.

The 't-test and 'chi-square' tests are essential tools in statistical analysis. The 't' test evaluates whether there is a significant difference between the means of two groups, which can be either related or independent. The chi-square test, on the other hand, assesses the association between categorical variables or tests the goodness of fit between observed and expected frequencies. Both tests help determine the significance of observed patterns and differences in the data.

Central tendency is a key concept in descriptive statistics, summarizing the centre point of a data distribution. It includes measures like the mean (the arithmetic average of a dataset), the median (the middle value when the data is ordered), and the mode (the most frequently occurring value). These measures provide a quick snapshot of the data's central value, guiding interpretations and decision-making processes based on sampled data. Understanding and applying these concepts are fundamental to effective data analysis and research.

2.7 GLOSSARY

- **Sampling Theory:** The field of statistics that deals with the selection of a subset (sample) from a larger population to estimate the characteristics of the entire population.
- **Probability:** The measure of the likelihood that a particular event will occur, expressed as a number between 0 and 1.
- **'t' Test:** A statistical test used to determine if there is a significant difference between the means of two groups. It is commonly used when the sample sizes are small and the population standard deviation is unknown.

- **Chi-Square Test:** A statistical test used to examine the association between categorical variables or to compare observed data with expected data according to a specific hypothesis.
- **Significance:** A statistical measure that indicates whether the results of an analysis are likely to be true and not due to random chance. It is often determined using a p-value.
- **Central Tendency:** A measure that represents the centre or typical value of a dataset. The main measures of central tendency are the mean, median, and mode.
- **Mean:** The arithmetic average of a set of values, calculated by summing the values and dividing by the number of values.
- **Median:** The middle value in a dataset when the values are arranged in ascending or descending order. If the dataset has an even number of observations, the median is the average of the two middle numbers.
- **Mode:** The value that appears most frequently in a dataset. A dataset may have one mode, more than one mode, or no mode at all.
- **P-value:** A statistical measure that helps determine the significance of the results. It represents the probability of obtaining results at least as extreme as the observed results, assuming the null hypothesis is true. A p-value less than a chosen significance level (e.g., 0.05) indicates strong evidence against the null hypothesis.

2.8 ANSWER TO THE CHECK YOUR PROGRESS

1. Which of the following is a measure of central tendency?

(a) Variance

(b) Range

(c) Median

(d) Standard deviation

Answer: C

2. What is the measure of central tendency that represents the middle value of a data set?

(a) Mean

(b) Mode

(c) Median

(d) Variance

Answer: C

3. If the mean of a data set is 10, and the data set has 5 values, what is the sum of all values in the data set?

(a) 10

(b) 20

(c) 50

(d) 100

Answer: D

4. What is the measure of dispersion that represents the average deviation of data points from the mean? (a) Standard deviation

(b) Range

(c) Variance

(d) Mode

Answer: A

5. If the range of a data set is 15 and the minimum value is 5, what is the maximum value?

(a) 10

(b) 15

(c) 20

(d) 25

Answer: B

6 Which statistical test is used to determine if there is a significant difference between the means of two groups?

(a) Chi-square test

(b) ANOVA

(c) t-Test

(d) Linear regression

Answer: C

7. What is the null hypothesis in a chi-square test?

(a) There is no association between the two variables.

(b) There is a significant association between the two variables.

(c) There is no significant difference between the means of the two groups.

(d) There is a significant difference between the means of the two groups.

Answer: D

8. ANOVA is used to determine if there is a significant difference in means among how many groups.

(a) Two

(b) Three or more

(c) Four

(d) Five or more

Answer: B

9. Which measure of central tendency is not affected by extreme values in the data set?

(a) Mean

(b) Median

(c) Mode

(d) Standard deviation

Answer: B

10. Which statistical test is appropriate for analyzing the relationship between two categorical variables?

(a) t-Test

(b) ANOVA

(c) Chi-square test

(d) Linear regression

Answer: C

2.9 REFERENCES

- https://mis.alagappauniversity.ac.in/siteAdmin/dde-admin/uploads/4/__PG_M.Com_Commerce%20(English)_Research%20Methodology_6223.pdf https://www.questionpro.com/blog/types-of-sampling-for-social-research/
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). Introduction to the Practice of Statistics. W.H. Freeman.
- Agresti, A. (2018). Statistical Methods for the Social Sciences. Pearson.
- Gravetter, F. J., & Wallnau, L. B. (2016). Statistics for the Behavioral Sciences. Cengage Learning.
- Triola, M. F. (2018). Elementary Statistics. Pearson.

2.10 TERMINAL QUESTIONS

1. What are the main uses of statistics in the field of business and economics?
2. Describe three measures of central tendency commonly used in statistical analysis.
3. Explain the concept of variance and its significance in statistics.
4. What is the purpose of conducting a hypothesis test in statistical analysis?
5. Discuss the difference between a t-test and an ANOVA test, and when each is appropriate to use.
6. How do you interpret the p-value obtained from a statistical test?
7. Describe the relationship between correlation and covariance.
8. In what scenarios would you use a chi-square test for independence?
9. Explain the importance of confidence intervals in inferential statistics.
10. What is the role of standard deviation in measuring the spread of data?
11. Discuss the limitations of using the mean as a measure of central tendency.
12. Compare and contrast the applications of parametric and non-parametric statistical tests.
13. What is the probability of rolling a prime number with a fair six-sided die?
14. A box contains 4 red balls, 3 green balls, and 5 blue balls. If you randomly select one ball, what is the probability of selecting a red or a blue ball?
15. If you flip a fair coin three times, what is the probability of getting exactly two tails?

UNIT-3 CORRELATION: KARL PEARSON & SPEARMAN'S REGRESSION ANALYSIS CARTOGRAPHY & REPRESENTATION OF SOCIO-ECONOMIC DATA: DOTS, PROPORTIONAL SQUARES, CIRCLES, SPHERES, CHOROPLETH, STAR & SCATTER DIAGRAMS, ERGOGRAPHS, AND FLOW LINE DIAGRAMS

3.1 OBJECTIVES

3.2 INTRODUCTION

3.3 CORRELATION: KARL PEARSON & SPEARMAN'S

3.4 SUMMARY

3.5 GLOSSARY

3.6 ANSWERS TO CHECK YOUR PROGRESS

3.7 REFERENCES

3.8 TERMINAL QUESTIONS

3.1 OBJECTIVES

After studying this chapter you will be able to:

- Learners will be able to construct dot maps, explaining their use in representing the distribution and density of socio-economic phenomena.
- Understand and explain the concept of Karl Pearson’s correlation coefficient, including its range and interpretation of values.
- Learners will be able calculate and interpret Pearson’s and Spearman’s correlation coefficients using given data sets.
- Students will gain the ability to create and understand proportional squares and circles as visual representations of data quantities, recognizing their proportional characteristics.
- Develop and analyze choropleth maps to show variations in socio-economic indicators within predefined areas, such as regions or countries.
- Generate and interpret star diagrams for comparing multiple socio-economic variables simultaneously.

3.2 INTRODUCTION

The analysis and representation of socio-economic data are crucial for understanding complex relationships and patterns within societies. Correlation analysis, pioneered by Karl Pearson and Charles Spearman, offers statistical methods to measure and interpret the strength and direction of relationships between variables. Pearson's correlation coefficient evaluates linear associations, while Spearman's rank correlation assesses monotonic relationships using ranked data. Complementing these statistical techniques, cartographic methods visually represent socio-economic data through various forms such as dot maps, proportional symbols (squares, circles, spheres), choropleth maps, star diagrams, scatter diagrams, ergographs, and flow line diagrams. These visual tools transform data into accessible and interpretable formats, aiding fields like geography, urban planning, and public policy in making informed decisions and revealing underlying trends.

3.3 CORRELATION: KARL PEARSON & SPEARMAN’S REGRESSION ANALYSIS CARTOGRAPHY & REPRESENTATION OF SOCIO-ECONOMIC DATA

Karl Pearson Coefficient

The Karl Pearson Coefficient of Correlation is an essential statistical tool in geography for measuring the strength and direction of the linear relationship between two continuous variables. This method quantifies the extent to which changes in one variable are associated with changes in another, which is crucial for analyzing and understanding geographical phenomena.

Also known as the Product Moment Correlation Coefficient, this method was developed by Karl Pearson. It is one of the three most powerful and widely used techniques for measuring correlation, alongside the Scatter Diagram and Spearman's Rank Correlation.

The Karl Pearson correlation coefficient method is quantitative, providing a numerical value that indicates the intensity of the linear relationship between X and Y. This coefficient is represented by 'r'.

Definition and Range

The Karl Pearson Coefficient of Correlation, denoted by 'r', ranges from -1 to +1:

a) $r=+1$: Perfect positive linear relationship.

b) $r=-1$: Perfect negative linear relationship.

c) $r=0$: No linear relationship.

What do You mean by Correlation Coefficient?

Before diving into the specifics of the Karl Pearson Coefficient of Correlation, it's essential to grasp the basic concepts of correlation and its coefficient. The correlation coefficient is a metric that quantifies the relationship between two variables, whether they are quantitative or qualitative, such as X and Y. This statistical tool is used to analyze and measure the strength and direction of their linear relationship.

For instance, changes in a person's monthly income (X) can affect their monthly expenditure (Y). By using the correlation coefficient, we can determine the extent to which variations in one variable influence the other.

Types of Correlation Coefficient

Correlation can be categorized into three types based on the direction of the relationship between variables:

i) Positive Correlation (0 to +1): This occurs when X and Y change in the same direction. For example, an increase in workout duration results in more calories burned.

ii) Negative Correlation (0 to -1): This occurs when X and Y change in opposite directions. For instance, an increase in the price of a commodity leads to a decrease in its demand.

iii) Zero Correlation (0): This indicates no relationship between the variables. For example, an increase in height does not affect one's intelligence.

With these basics in mind, let's proceed to the Karl Pearson Coefficient of Correlation.

Application in Geography

In geography, this coefficient helps in understanding and quantifying relationships between various physical and human geographical variables. Here are some key examples:

1.Climate and Agriculture

Example: Investigating the relationship between annual rainfall (X) and crop yield (Y).

Use: Helps in understanding how variations in rainfall influence agricultural productivity. A high positive r indicates that higher rainfall is associated with higher crop yields.

2.Urbanization and Pollution

Example: Studying the correlation between the level of urbanization (X) and air pollution levels (Y).

Use: A positive correlation can indicate that as cities grow, pollution levels tend to increase, guiding urban planning and environmental policies.

3.Topography and Erosion

Example: Analyzing the relationship between slope gradient (X) and soil erosion rates (Y).

Use: Helps in understanding how steep slopes are more susceptible to erosion, aiding in land management and conservation efforts.

Formula: The formula for calculating the Pearson correlation coefficient r is:

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

Where,

n = number of pairs of scores

X and Y are the variables

∑ denotes the sum of the values

Steps for Calculation

i)Collect Data: Gather paired data for the two variables X and Y.

ii) Compute Summations: Calculate the sums $\sum X$, $\sum Y$, $\sum XY$, $\sum X^2$ and $\sum Y^2$.

iii) Apply Formula: Substitute the calculated values into the Pearson correlation formula

1. **Example :**Consider a study examining the relationship between average annual rainfall (in mm) and crop yield (in tons per hectare) across different regions.

Table No. 3.1

Region	Annual Rainfall (X)	Crop Yield (Y)
A	800	2.5
B	600	2.0
C	1000	3.0
D	1200	3.5
E	900	2.8

Steps to Calculate the Pearson Correlation Coefficient

Data Collection: Gather data for rainfall and crop yield.

Compute Summations:

$\sum X$=800+600+1000+1200+900=4500

$\sum Y$=2.5+2.0+3.0+3.5+2.8=13.8

∑XY=(800×2.5)+(600×2.0)+(1000×3.0)+(1200×3.5)+(900×2.8)=2000+1200+3000+4200+2520=12920

$\sum X^2$=8002+6002+10002+12002+9002=640000+360000+1000000+1440000+810000=4250000

$\sum Y^2$=2.52+2.02+3.02+3.52+2.82=6.25+4.0+9.0+12.25+7.84=39.34

Substitute into Formula:

$$r=\frac{5(12920)-(4500)(13.8)}{\sqrt{[5(4250000)-(4500)^2][5(39.34)-(13.8)^2]}}$$

$$r=\frac{64600-62100}{\sqrt{[21250000-20250000][196.7-190.44]}}$$

$$r=\frac{2500}{\sqrt{1000000\times 6.26}}$$

$$r=2500\frac{2500}{2500}=1.0$$

This calculation shows a perfect positive linear relationship between annual rainfall and crop yield in the given regions.

2. A researcher is investigating the relationship between the level of urbanization (measured as the percentage of urbanized land) and the air pollution levels (measured as the concentration of PM2.5 in µg/m³) in different cities. The following data is collected:

Table No. 3.2

City	Urbanization Level (%)	Air Pollution Level (Y) (PM2.5 µg/m³)
1	30	35
2	45	50
3	60	65
4	75	80
5	90	95

Using the Karl Pearson Coefficient of Correlation, calculate the correlation coefficient r to determine the strength and direction of the relationship between the level of urbanization and air pollution levels. Interpret the result.

Answer: Calculate Summations:

$\sum X=30+45+60+75+90=300$

$\sum Y=35+50+65+80+95=325$

$\sum XY= (30\times35) + (45\times50)+(60\times65)+(75\times80)+(90\times95)=1050+2250+3900+ 6000+8550=21750$

$\sum X^2=302+452+602+752+902=900+2025+3600+5625+8100=20250$

$\sum Y^2=352+502+652+802+952=1225+2500+4225+6400+9025=23375$

Substitute into Pearson Formula:

$$r=\frac{n(\sum XY)-(\sum X)(\sum Y)}{\sqrt{[n\sum X^2-(\sum X)^2][n\sum Y^2-(\sum Y)^2]}}$$

Where n=5

$$r=\frac{5(21750)-(300)(325)}{\sqrt{[5(2025)-(300)^2][5(23375)-(325)^2]}}$$

$$r=\frac{108750-97500}{\sqrt{[101250-90000][116875-105625]}}$$

$$r=\frac{11250}{\sqrt{11250\times11250}}$$

$$r=\frac{11250}{11250}=1$$

Table No. 3.3

Correlation	Positive Value of r	Negative Value of r
Perfect Correlation	+ 1	-1
Very High Degree Correlation	Between .9 to .99	Between -.99 to -.9
High Degree Correlation	Between .75 and .9	Between -.9 to -.75
Moderate Degree Correlation	Between .25 and .75	Between -.75 to -.25
Low Degree Correlation	Between 0 and .25	Between -.25 and 0
No Correlation	0	0

Spearsman Regression Analysis

When the variables in question cannot be measured quantitatively but can be arranged in a sequential order (ranks), we determine the correlation between the ranks of two series. This scenario arises when dealing with qualitative traits like honesty, beauty, etc. This method, known as Spearman's Rank Difference Method or Ranking Method, yields a correlation coefficient called the Rank Correlation Coefficient, denoted by r_s. This technique was developed by British psychologist Charles Edward Spearman in 1904.

Spearman's rank correlation coefficient r_s. is also applicable when measurements are provided for both series. Essentially, r_s is equivalent to Karl Pearson's correlation coefficient but applied to the ranks, and it can be interpreted similarly to Pearson's correlation coefficient.

Spearman's rank correlation coefficient is a statistical tool used to measure the strength and direction of association between two variables in geography. Unlike Pearson correlation, which assumes a linear relationship between variables, Spearman correlation uses the ranks of data points rather than their actual values, making it suitable for analyzing non-linear relationships and ordinal data.

Spearman's Rank Correlation Coefficient

Correlation is a statistical measure that assesses how closely two variables change together. A positive correlation indicates the degree to which these variables increase or decrease in tandem. Conversely, a negative correlation indicates the extent to which one variable increases as the other decreases. Here we will focus on one type of correlation: Spearman's Rank Correlation.

What Is a Monotonic Function?

a) Monotonically Increasing: As the variable X increases, the variable Y does not decrease.

b) Monotonically Decreasing: As the variable X increases, the variable Y does not increase.

c) Not Monotonic: As the variable X increases, the variable Y sometimes increases and sometimes decreases.

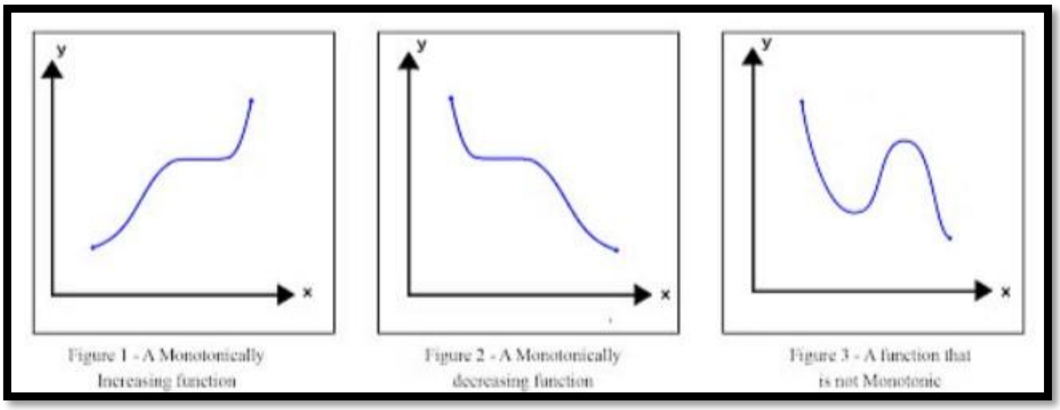


Fig 3.1: Monotonic Function, Source: Google

Spearman's rank correlation measures the strength and direction of association between two ranked variables. It basically gives the measure of monotonicity of the relation between two variables i.e. how well the relationship between two variables could be represented using a monotonic function.

Spearman's rank correlation assesses the strength and direction of the relationship between two ranked variables. It essentially measures the monotonicity of the relationship between the variables, indicating how well this relationship can be represented by a monotonic function.

The formula for Spearman's rank coefficient is:

$$\rho = 1 - \frac{6 \in d^2 i}{n(n^2 - 1)}$$

Where,

ρ= Spearman's rank correlation coefficient

d= Difference between the two ranks of each observation

n= Number of observations

Spearman's Rank Correlation Coefficient, denoted as r, has several distinct characteristics that make it a valuable tool in statistical analysis, especially when dealing with ordinal data or non-linear relationships. Here are the key characteristics:

1. Non-parametric Measure: Spearman's rank correlation is a non-parametric measure, meaning it does not assume any specific distribution for the data. It can be used with ordinal data or when the assumptions for Pearson's correlation are not met.

2. Based on Ranks: It evaluates the relationship between variables using the ranks of the data points rather than their actual values. This makes it robust against outliers and non-linear relationships.

3. Range: The coefficient ranges from -1 to +1:

ρ=+1 indicates a perfect positive monotonic relationship.

ρ=−1 indicates a perfect negative monotonic relationship.

ρ=0 indicates no monotonic relationship.

4. Monotonic Relationship: It assesses monotonic relationships, which means that as one variable increases, the other variable either consistently increases or decreases, but not necessarily at a constant rate.

5. Handling Ties: Spearman's rank correlation can handle tied ranks by assigning them the average of the ranks they would have received if there were no ties.

6.Applicability: Suitable for both small and large datasets. It is particularly useful when the data is ordinal, or the relationship between variables is monotonic but not linear.

Applications in Geography

1. Climate Studies:

- **Temperature and Altitude:** Spearman's rank correlation can be used to assess the relationship between temperature and altitude. By ranking various locations by altitude and average temperature, researchers can determine if higher altitudes generally correspond to lower temperatures.
- **Precipitation Patterns:** Analyzing the correlation between rainfall amounts and geographical factors such as distance from the sea or elevation.

2. Population and Urban Studies:

- **Urbanization and Population Density:** Assessing the correlation between the degree of urbanization and population density in different areas.
- **Infrastructure and Economic Development:** Examining the relationship between the development of infrastructure (like roads and public services) and economic indicators such as GDP or employment rates in different regions.

3. Environmental Studies:

- **Deforestation and Land Use:** Evaluating the relationship between deforestation rates and land use practices, such as agriculture or urban expansion.
- **Pollution and Health:** Correlating pollution levels with health outcomes across different regions to identify areas where pollution might be impacting public health.

4. Hydrology and Water Resources:

- **River Discharge and Rainfall:** Analyzing the correlation between river discharge rates and rainfall amounts in different catchment areas.
- **Water Quality and Land Use:** Studying how different land uses (industrial, agricultural, residential) correlate with water quality indicators in nearby water bodies.

Example Calculation in a Geographical Context

1.Consider a study examining the correlation between the level of urbanization (X) and air quality index (Y) in different cities. The data might look like this:

Table No. 3.4

City	Urbanization Rank(X)	Air Quality Rank (Y)
City A	1	3
City B	2	2
City C	3	1
City D	4	5
City E	5	4

3. Compute d_i and d^2_i

Table No. 3.5

Rank (X)	Rank (Y)	d_i (X-Y)	d^2_i
1	3	-2	4
2	2	0	0
3	1	1	1
4	5	-1	-1
5	4	1	1

3. d^2_i: 4+ 0+4+1+1=10

4. Apply the Formula: $\rho = 1 - \frac{6 \in d^2 i}{n(n^2-1)}$

Apply the formula,

$$\rho = 1 - \frac{6 \times 10}{5(5^2 - 1)}$$

$$\rho = 1 - \frac{60}{5(25 - 1)}$$

$$\rho = 1 - \frac{60}{120}$$

$$\rho = 1 - 0.5$$

$$\rho = 0.5$$

So, ρ=0.5, indicating a moderate positive correlation between urbanization level and air quality rank in this example.

Interpretation in Geographical Studies

a) Positive Correlation: A positive Spearman's rank correlation coefficient suggests that as one variable increases, the other variable tends to increase as well. For example, higher urbanization might be correlated with higher pollution levels.

b) Negative Correlation: A negative coefficient indicates that as one variable increases, the other tends to decrease. For instance, higher altitudes might correlate with lower temperatures.

c) Zero Correlation: A coefficient close to zero suggests no monotonic relationship between the variables, meaning changes in one variable do not predict changes in the other.

2. Consider the following data set of two variables, X and Y, collected from five different locations:

Table No. 3.6

Location	X	Y
A	85	70
B	60	80
C	75	65
D	90	85
E	70	75

1. Rank the data for both variables X and Y.

2. Calculate the differences (d) between the ranks of each pair of observations.

3. Compute Spearman's rank correlation coefficient (ρ) using the formula

$$\rho = 1 - \frac{6 \in d^2 i}{n(n^2 - 1)}$$

4. Interpret the result of the correlation coefficient.

Solution:

1. **Rank the Data:**

Table No. 3.7

Location	X	Rank(X)	Y	Rank (Y)
A	85	4	70	2
B	60	1	80	4
C	75	3	65	1
D	90	5	85	5
E	70	2	75	3

2. **Calculate the Differences and Squared Differences:**

Table No. 3.8

Location	Rank(X)	Rank (Y)	d_i (X-Y)	d^2_i
A	4	2	2	4
B	1	4	-3	9
C	3	1	2	4
D	5	5	0	0
E	2	3	-1	1

3. Sum of Squared Differences: $\sum d_i$=4+9+4+0+1=18

4. Calculate Spearman's Rank Correlation Coefficient:

$$\rho = 1 - \frac{6 \in d^2 i}{n(n^2 - 1)}$$

$$\rho = 1 - \frac{6 \times 18}{5(5^2 - 1)}$$

$$\rho = 1 - \frac{108}{120}$$

$$\rho = 1 - 0.9$$

$$\rho = 0.1$$

Interpretation: The Spearman's rank correlation coefficient (ρ/ρ) of 0.1 indicates a very weak positive correlation between the variables X and Y. This means that there is a slight tendency for higher values of X to be associated with higher values of Y, but the relationship is not strong.

Dots, Proportional squares, Circles, Spheres, Choropleth, Star & Scatter diagrams, Ergographs, Flow line diagrams.

A thematic map is a type of map that visualizes specific spatial patterns, distributions, or relationships related to a particular theme or topic. Unlike general reference maps, which primarily display geographic features such as roads, rivers, and political boundaries, thematic maps focus on conveying specific information about a chosen subject matter. Thematic maps use various graphical techniques, such as colour coding, symbols, and shading, to represent quantitative or qualitative data associated with the chosen theme. These maps are widely used in disciplines such as geography, environmental science, urban planning, sociology, and economics to analyze and communicate spatial patterns and trends.

Some of the examples of thematic maps discussed in this unit include dot maps, proportional symbol maps using squares, circles, and spheres, choropleth maps, star and scatter diagrams, ergographs, and flow line diagrams. These thematic maps employ various graphical representations to visualize specific spatial patterns, distributions, or relationships related to a particular theme or topic. Now we discuss about them one by one:

Dot Map

In this thematic mapping technique, the identical dots are used to explain the occurrence of a particular geographic theme like distribution and density of population and livestock resources. It provides a quantitative value with the help of a plotted dots for the visual representation. However, such dots placed on a map are rarely used for quantitative interpretation. Hence, dot maps are different than choropleth maps as they cannot provide the idea of area and quantitative measurement as such. Here, few important ones are discussed out of the several methods as under.

A dot map is a type of thematic map that represents spatial phenomena using individual dots to symbolize discrete geographic entities or data points. Each dot on the map typically represents a fixed quantity or a specific attribute associated with the chosen theme or topic. Dot maps are particularly useful for visualizing point data, such as the locations of cities, landmarks, or events, as well as for illustrating spatial patterns or distributions within a given area.

Components of a Dot Map:

a) Dots: The primary visual element of a dot map, each dot represents a specific geographic entity or data point. The size, colour, or shape of the dots may vary based on the data being represented.

b) Base Map: Dot maps are usually overlaid on top of a base map, which provides context and reference information such as geographic features, roads, and political boundaries. The base map helps viewers orient themselves spatially and understand the geographic context of the dot distribution.

c) Legend: A legend accompanies the dot map to provide interpretation of the dots' meaning. It explains the relationship between the size, colour, or other attributes of the dots and the corresponding data values or categories.

Characteristics and Usage:

a) Point Data Representation: Dot maps are especially effective for representing point data, where each dot corresponds to a specific geographic location or point of interest.

b) Quantitative and Qualitative Data: Dot maps can represent both quantitative and qualitative data. For quantitative data, the number of dots in a particular area may represent a count or density. For qualitative data, dots can be coloured or shaped differently to represent different categories or attributes.

c) Spatial Patterns and Distributions: Dot maps help visualize spatial patterns and distributions within a geographic area. They allow viewers to identify clusters, concentrations, or gaps in the distribution of the represented data.

d) Scale Considerations: The scale of a dot map is crucial for accurate interpretation. Large-scale dot maps may show individual data points with high detail, while small-scale dot maps may aggregate data into broader patterns over larger geographic areas.

Types of dot map:

a) Mono-dot-Map: The spatial distribution of various socio-economic variables can be effectively portrayed using the dot mapping technique. This method involves placing dots of uniform size on a map to convey relative density, offering a visual impression without providing precise numerical values. Counting the dots on the map is required to obtain absolute figures. Dot maps are relatively straightforward to create compared to other mapping techniques, as they do not require complex computations or manual skills. Typically, mono-dot maps depict the distribution of a single geographical variable, such as crop acreage, irrigation facilities, or population density. However, multi-coloured dot maps can represent multiple variables simultaneously; for instance, male and female literacy rates can be depicted in different colours on a literacy map. Care must be taken to select appropriate dot sizes to prevent overcrowding or sparsity on the map. Additionally, spacing between dots within each geographical unit should be uniform, especially for larger administrative units, to minimize visual and spacing discrepancies.

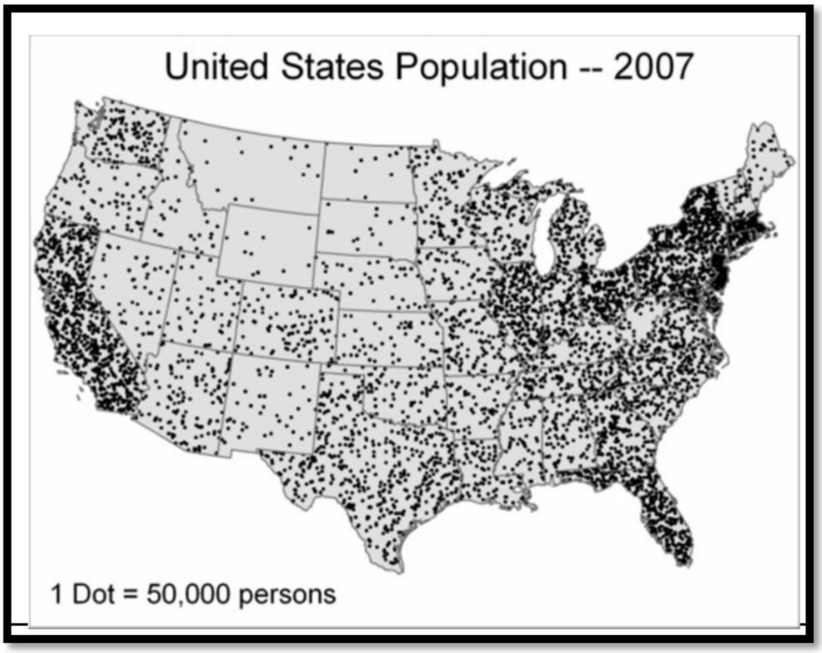


Fig 3.2: Mono-dot Map, Source: Google

Dot density maps utilize dots to signify specific quantities. In the map shown, each dot stands for 50,000 people. These dots reflect the population within designated enumeration units, such as states. The greater the number of dots, the larger the population—hence, California has more dots than Arizona due to its higher population.

b) Multiple Dot Map: A multiple dot map in cartography is a thematic map that employs various types or sizes of dots to represent different attributes, categories, or levels of a phenomenon or variable across a geographic area. Unlike mono dot maps, which use a single type of dot to represent the distribution or density of a single variable, multiple dot maps allow for the simultaneous visualization of multiple variables or subgroups within a dataset.

In a multiple dot map, each type or size of dot may correspond to a specific range of values, categories, or attributes within the dataset. For example, different colours, shapes, or sizes of dots may represent different levels of income, population density, or education attainment in a given area. By using multiple types of dots, the map can convey more nuanced information and facilitate the comparison of different subgroups or characteristics within the dataset.

Multiple dot maps are valuable tools for visualizing complex spatial patterns, relationships, and distributions within datasets that involve multiple variables or attributes. They enable cartographers, researchers, and analysts to identify spatial correlations, disparities, and trends across various dimensions of the data, thereby enhancing understanding and insight into geographic phenomena.

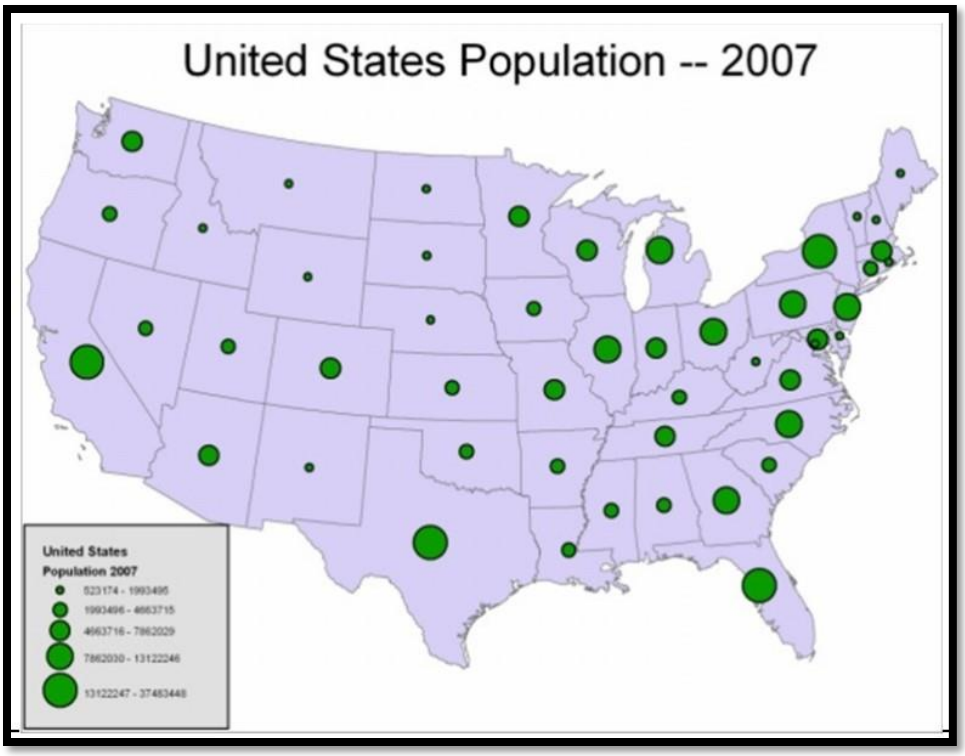


Fig 3.3: Multiple dot Map, Source: Google

Graduated symbols represent similar data to dot density maps but use varying symbol sizes within the enumeration areas. In this type of map, the larger the symbol, the greater the represented quantity. Thus, in this map, California has a larger symbol than Arizona, reflecting its larger population.

Examples of Dot Map Applications:

a) Population Distribution: Dot maps can illustrate the distribution of population by placing a dot for each individual or for a certain number of people within a specific area.

b) Crime Mapping: Law enforcement agencies use dot maps to visualize the locations of criminal incidents, enabling them to identify hotspots and allocate resources effectively.

c) Epidemiology: Dot maps are used in epidemiological studies to track the spread of diseases by mapping the locations of reported cases.

d) Retail Analysis: Businesses use dot maps to visualize the locations of customers, stores, or competitors to inform marketing strategies and site selection for new outlets.

Advantages:

1) **Visual Representation:** Dot maps provide a clear and intuitive visual representation of spatial distributions and patterns, making it easy for viewers to interpret the data.
2) **Relative Density:** By varying the density or clustering of dots, dot maps can convey relative levels of intensity or concentration of the phenomenon being mapped.
3) **Comparative Analysis:** Dot maps allow for the comparison of spatial patterns and distributions across different regions or areas of interest, enabling comparative analysis.

4) **Simple Construction:** Dot maps are relatively easy to construct and interpret compared to other mapping techniques, as they do not require complex computations or manual skills.
5) **Multivariate Representation:** Multiple dot maps can represent multiple variables or attributes simultaneously, providing insights into complex spatial relationships and interactions.

Disadvantages:

1) **Lack of Absolute Values:** Dot maps do not provide absolute numerical values for the data being represented, requiring viewers to count the dots to obtain precise figures.
2) **Scale Sensitivity:** The interpretation of dot maps can be sensitive to changes in scale, as variations in dot density may appear differently at different scales.
3) **Subjectivity in Symbolization:** The selection of dot size, shape, and color can introduce subjectivity and bias into the map interpretation, potentially affecting the accuracy of the representation.
4) **Potential Overcrowding:** If not carefully managed, dot maps can suffer from overcrowding, where dense clusters of dots obscure underlying patterns or make the map difficult to read.
5) **Limited Detail:** Dot maps may lack detail compared to other mapping techniques, particularly for datasets with fine-grained spatial variations or complex patterns.

Proportional symbol map: Proportional square, Proportional circle, Proportional sphere

Proportional symbol maps are a type of thematic map used in geography to visualize spatial patterns and variations in data. They are designed to represent quantities or magnitudes of a particular variable across different geographic areas. These maps use symbols, typically varying in size, to indicate the magnitude of the variable being mapped at specific locations.

Proportional Square:

A proportional square diagram in cartography is a thematic map technique where square symbols of varying sizes are used to represent data quantities within geographical areas. The size of each square is proportional to the value it represents, allowing for a visual comparison of the magnitude of different data points across the map. This method helps to effectively communicate variations in data, such as population or economic metrics, across different regions.

Here's a detailed explanation of proportional square maps:

1. Variable Selection: Similar to other thematic maps, the first step in creating a proportional square map is selecting the variable you want to visualize spatially. This variable could be anything quantitative that varies across geographic regions, such as population density, income levels, land area, or any other relevant metric.

2. Data Collection: Once the variable is chosen, data needs to be collected for the geographic areas of interest. This data could come from governmental sources, research institutions, surveys, or other reliable sources. It's essential that the data is accurate, up-to-date, and consistent across the geographic regions being mapped.

3. Square Design: Instead of using circles or other symbols, proportional square maps utilize squares to represent data. The size of each square corresponds to the value of the variable being mapped. Larger squares represent higher values, while smaller squares represent lower values. The design of the squares should be visually appealing and easily distinguishable from one another.

4.Scaling: Scaling is crucial in proportional square maps to ensure that the size of each square accurately reflects the magnitude of the variable being mapped. The scaling can be linear or logarithmic, depending on the range and distribution of the data. It's important to choose a scaling method that effectively communicates the relative differences in the variable across geographic areas.

5.Arrangement: Once the squares are designed and scaled, they need to be arranged on the map at the appropriate geographic locations. The arrangement should accurately represent the spatial distribution of the variable being mapped. Squares may overlap in densely populated or high-value areas, so careful consideration is needed to ensure clarity and readability of the map.

6. Coloration and Labelling: Colours can be used to enhance the visual representation of the map and differentiate between different categories or ranges of values. For example, a colour gradient ranging from light to dark could be used to represent increasing values of the variable. Labels may also be added to the squares or placed adjacent to them to provide additional context and information about the data being mapped.

7. Interpretation: Once the proportional square map is created, it needs to be interpreted to extract meaningful insights about the spatial patterns and variations in the data. Viewers can analyze the map to identify areas with high or low values of the variable, detect spatial trends or clusters, and make comparisons between different geographic regions.

Proportional square maps are valuable tools for visualizing quantitative spatial data and gaining insights into geographic patterns and relationships. They are commonly used in fields such as urban planning, demographic analysis, resource management, and environmental science.

Depending on the nature of data square diagram may be constructed in two ways:-

(i) Simple square diagram or proportional square diagrams where data with only one component is available and each square represent a single value.

(ii) Compound square diagrams when the comprise more than one component and each square is proportionally divided into rectangular segment to show the construction parts.

Principal: The construction of square diagram is based on the principal that the are of square(a) for a responsible item (i) directly proportional to the quantity of the item (q) to be shown.

Method: Square diagram is drawn by following the step: One side of square is simply the square root of the data $L^2 = x$ (where ‘x’ number of household) and house. Or $L=\sqrt{x}$

Use: Square usually represents the number of household and house etc.

Proportional Circle Map:

What is a Proportional Circle Map?

A proportional circle map is a type of thematic map used in geography to represent quantitative data visually. Unlike traditional maps that show geographical features like rivers and mountains, thematic maps focus on specific themes or topics, such as population density, income distribution, or climate patterns.

How Does it Work?

In a proportional circle map, circles of varying sizes are used to represent different quantities or magnitudes of a particular variable across geographic areas. Larger circles represent higher values of the variable, while smaller circles represent lower values. For example, if you're mapping population density, larger circles might represent densely populated areas, while smaller circles represent sparsely populated regions.

Creating a Proportional Circle Map:

Choose a Variable: Decide on the variable you want to visualize spatially. This could be population, income, temperature, or any other quantitative data.

Collect Data: Gather data for the chosen variable for the geographic areas you're interested in. Ensure the data is accurate and up-to-date.

Design the Circles: Design circles to represent the data. The size of each circle should correspond to the value of the variable being mapped. Larger circles indicate higher values, while smaller circles indicate lower values.

Scale the Circles: Scale the circles appropriately to accurately reflect the magnitude of the variable. You can use linear or logarithmic scaling, depending on the range and distribution of the data.

Place the Circles: Position the circles on the map according to their geographic locations. Ensure that the placement accurately represents the spatial distribution of the variable.

Colour and Label: Use colours to enhance the visual representation of the map. You can use a colour gradient to differentiate between different ranges of values. Labels can also be added to provide additional context and information.

Interpreting the Map:

Once the proportional circle map is created, it can be interpreted to understand spatial patterns and variations in the data. Look for areas with larger circles, as they indicate higher values of the variable. Conversely, smaller circles represent lower values. Analyze the map to identify trends, clusters, and relationships between different geographic regions.

Proportional circle maps are powerful tools for visualizing quantitative spatial data and gaining insights into geographic patterns. By understanding how to create and interpret these maps, distance learning students can develop a deeper understanding of geographical concepts and phenomena.

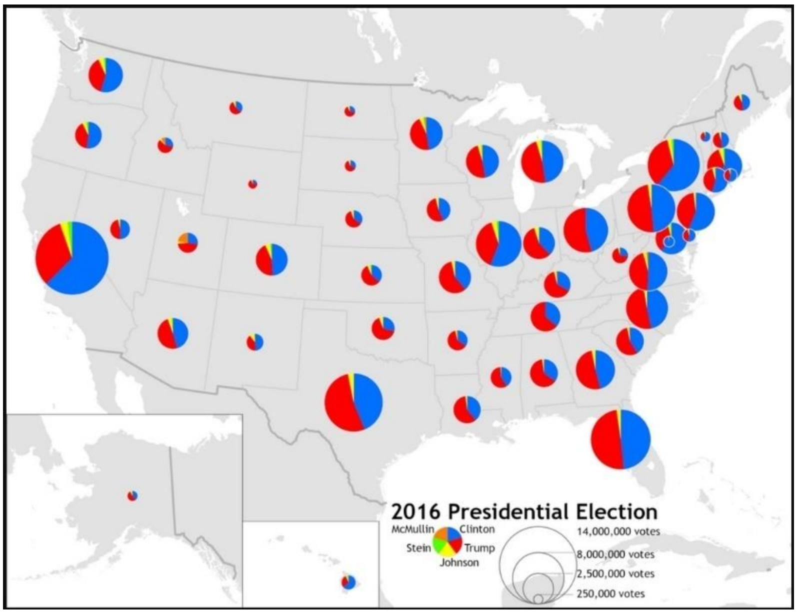


Fig3.4: Proportional Circle Map, Source: Google

A proportional symbol map of the 2016 U.S. presidential election, in which the circles are proportional to the total number of votes cast in each state, formatted as a pie chart showing the relative proportion for each candidate.

Proportional Sphere Map:

A proportional sphere map, also known as a three-dimensional proportional symbol map, is a specialized type of thematic map used in geography to represent quantitative data in three-dimensional space. Unlike traditional maps that are flat representations of the Earth's surface, proportional sphere maps use spheres of varying sizes to visualize data across different geographic regions. Here's a detailed explanation of proportional sphere maps:

1. Conceptual Basis:

- Proportional sphere maps are based on the principle of proportional symbol mapping, where symbols' sizes are proportional to the quantitative variable being represented.
- In this case, instead of using two-dimensional symbols like circles or squares, spheres are employed to represent the data in three-dimensional space.

2. Variable Selection and Data Collection:

- As with other thematic maps, the first step is to select the variable of interest, such as population, GDP, or pollution levels.
- Data for this variable is then collected for the geographic areas of interest, ensuring accuracy and consistency across regions.

3. Sphere Design and Scaling:

- Spheres are designed to represent the data, with their sizes indicating the magnitude of the variable.
- Scaling of the spheres is crucial to accurately reflect the variable's values. This can be done using linear or logarithmic scaling methods, depending on the data's range and distribution.

4. Placement in Three-Dimensional Space:

- Unlike traditional maps where symbols are placed on a flat surface, spheres in proportional sphere maps are positioned in three-dimensional space.
- The placement of spheres should accurately reflect the spatial distribution of the variable across geographic regions.

5. Coloration and Labelling:

- Colours can be used to enhance the visual representation of the map, helping differentiate between different categories or ranges of values.
- Labels may also be added to the spheres or placed adjacent to them to provide additional context and information about the data being mapped.

6. Interpretation:

- Once the proportional sphere map is created, it can be interpreted to extract insights into spatial patterns and variations in the data.
- Viewers can analyze the map to identify areas with larger spheres, indicating higher values of the variable, and compare them with areas containing smaller spheres representing lower values.

Applications: Proportional sphere maps are particularly useful for visualizing spatial data in fields such as urban planning, environmental science, and geology, where three-dimensional representations can provide additional insights into geographic phenomena.

Choropleth Map

A choropleth map is a thematic map in which areas are shaded or patterned proportionally to the value of a particular variable measured for each area. Most often the variable is quantitative, with a colour associated with an attribute value. Though not as common, it is possible to create a choropleth map with nominal data. Choropleth maps illustrate the value of a variable across the landscape with colour that changes across the landscape within a particular geographic area. A choropleth map is an excellent way to visualize how a measurement varies across a geographic area.

In very simple language we can say that 'A choropleth map is a specialized type of thematic map that effectively illustrates geographic regions by using distinct colours, shades, or patterns to represent different values of a chosen variable'.

The earliest known choropleth map was created in 1826 by Baron Pierre Charles Dupin.

Components of a Choropleth Map

a) **Geographic Areas:** The map is divided into distinct geographic units, such as countries, states, counties, or other administrative regions.

b) **Data Classification:** Data are aggregated for each geographic unit. This could be absolute values (like population count) or rates/ratios (like population density or average income).

c) **Shading or Colouring:** Each geographic unit is shaded or coloured according to the value of the statistical variable it represents. Typically, a gradient of colours is used, where lighter shades represent lower values and darker shades represent higher values.

For example, this map shows the population density in the year 2007 for the United States of America. For each state, the number of persons per square mile has been calculated. The states with a lower population density are shaded with a lighter gray colour. The states of a higher population density are shaded with a dark gray colour. The states with population densities between the two extremes are shaded on a continuum from the lightest gray to the darkest gray. Based on the five different shades of Gray, the map visually represents in an intuitive manner where the most densely populated states are in the United States.

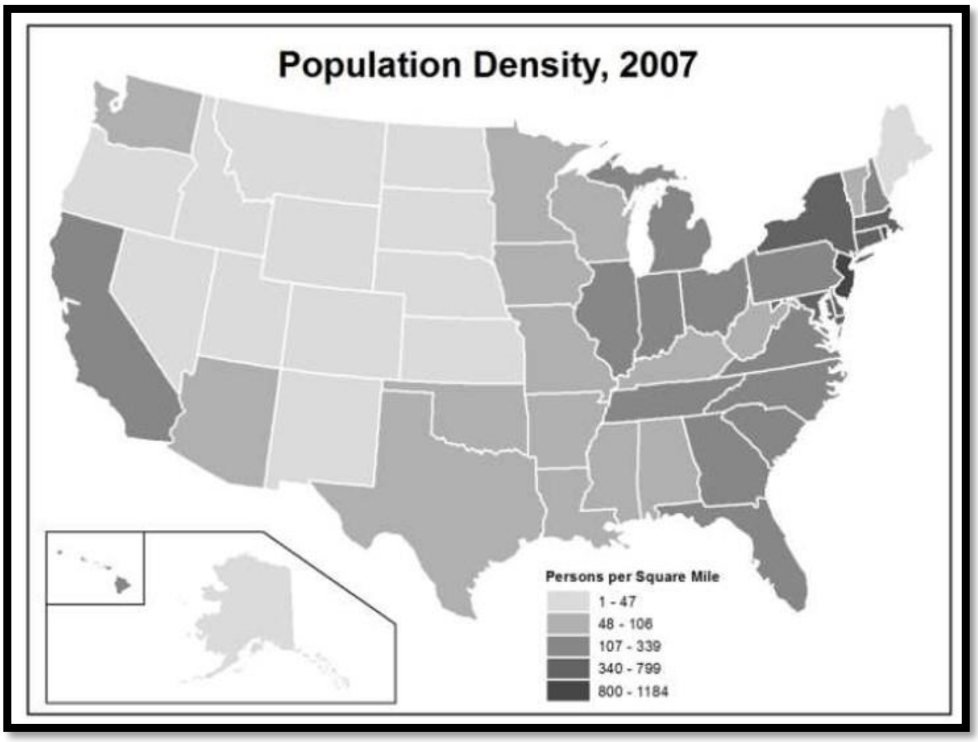


Fig3.5: Choropleth Map, Source: Google

Star & scatter Diagram

Wind Rose or Star Diagram

In cartography there are mainly two type of star diagram, i.e. wind rose diagram and simple star diagram. These are a graphical tool used to display the frequency and direction of wind or other directional data at a particular location over a specified period. It is particularly useful in meteorology, oceanography, and environmental studies for visualizing wind patterns.

Star diagrams are mainly of the following types:

i) Wind Rose Diagram

ii) Simple Star Diagram

i) Wind Rose Diagram: The most common star diagram is a wind rose diagram. This diagram is known as a star diagram. As you know, there are sixteen directions – four cardinal or primary directions and twelve secondary directions. Therefore, in the wind rose, a maximum of sixteen lines can be drawn from the center representing the corresponding sixteen directions. The length of each line would be proportionate to the quantity it means.

So, each ray will represent the number of hours or days the wind blows from the corresponding direction in a particular period. But there are some hours or days (as the case may be) when the wind is calm. These calm periods are generally shown by drawing a small circle at the center and writing the number within the circle. After all the lines are drawn, the endpoints of all the lines are joined.

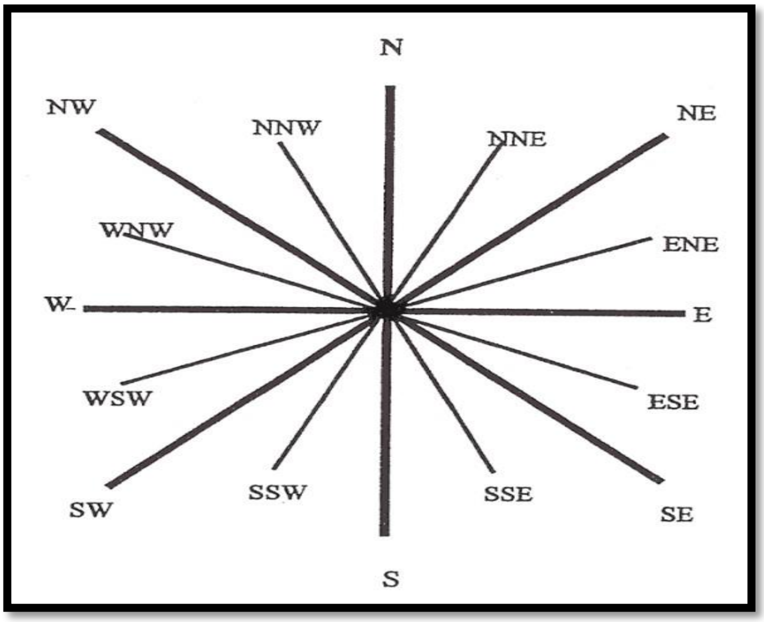


Fig 3.6: Four cardinal and twelve secondary directions, Source: Google

A wind rose diagram, commonly used in geography and meteorology, represents the frequency and direction of winds at a specific location over a given period of time. Here's a detailed explanation of a wind rose diagram:

a) Circular Layout: A wind rose diagram is typically presented in a circular format. The circle is divided into sectors or segments, each representing a specific compass direction (e.g., north, northeast, east, etc.).

b) Radial Axes: Radial lines extend from the centre of the circle to the outer edge, dividing the circle into equal segments. These lines represent the frequency or percentage of time that the wind blows from each direction.

c) Rose Petals: Each segment of the circle is filled with "petals" or "bars" that extend outward from the centre. The length of these petals corresponds to the frequency or proportion of time that the wind blows from the corresponding direction.

d) Colour or Shading: Wind roses often use colour or shading to indicate wind speed ranges within each direction sector. For example, darker shades may represent stronger winds, while lighter shades represent lighter winds.

e) Interpretation: By examining the wind rose diagram, one can identify prevailing wind directions (the directions with the longest petals), as well as secondary wind patterns. Additionally, the distribution of petal lengths provides insights into wind variability and intensity.

Applications: Wind rose diagrams are valuable tools for various applications, including climate analysis, urban planning, renewable energy assessment (such as for wind farms), and air quality management. They help in understanding local wind patterns, which in turn influence factors like temperature, precipitation, and air pollution dispersion.

Data Collection: To construct a wind rose diagram, meteorological data is collected over a specified period, typically using instruments like anemometers placed at a weather station. The data is then processed to calculate wind direction frequencies and speeds for each direction sector.

Overall, wind rose diagrams provide a visual representation of wind patterns, aiding in the analysis and understanding of regional climate dynamics and environmental processes.

Creating a wind rose diagram involves several steps, from collecting data to designing the visual representation. Here's a guide to creating a wind rose diagram in geography:

i) Collect Wind Data: Obtain wind data for your desired location from a reliable source such as a meteorological station or online weather databases. Data should include wind direction and wind speed measurements recorded at regular intervals (e.g., hourly or daily) over a specific period, such as a month or a year.

ii) Organize Data: Organize the collected wind data into a tabular format. Typically, the table will have columns for wind direction (in compass degrees) and wind speed (in units like meters per second or miles per hour) for each time interval.

iii) Calculate Frequency: Calculate the frequency or percentage of time that the wind blows from each direction. Count the number of occurrences of wind from each direction and divide by the total number of observations. This will give you the proportion of time the wind blows from each direction.

iii) Group Wind Directions: Group wind directions into directional sectors to simplify the presentation. For example, you might group directions into eight or sixteen sectors (e.g., N, NE, E, SE, S, SW, W, NW). Sum the frequencies of wind observations within each sector.

iv) Determine Wind Speed Ranges: Decide on wind speed ranges to represent in the wind rose diagram. Commonly used ranges include calm (0-1 m/s), light breeze (1-3 m/s), moderate breeze (3-6 m/s), strong breeze (6-9 m/s), and so on.

v) Calculate Wind Speed Frequencies: Calculate the frequency or percentage of time that winds fall within each speed range for each direction sector. This can be done by counting the number of observations within each range for each sector and dividing by the total number of observations in that sector.

Design the Wind Rose Diagram: Using graphing software or specialized wind rose plotting tools, create a circular diagram with radial axes representing wind direction sectors and concentric rings representing wind speed ranges. Plot the wind frequency or percentage for each direction sector and wind speed range as petals or bars extending from the center of the diagram.

Label and Annotate: Label the wind direction sectors and wind speed ranges on the wind rose diagram. Include a legend to explain the color or shading scheme used to represent wind speed ranges. Add any necessary annotations or explanations to aid interpretation.

Review and Interpret: Review the completed wind rose diagram to ensure accuracy and clarity. Interpret the diagram to identify prevailing wind directions, wind speed distribution, and any notable patterns or trends in the data.

Optional Enhancements: Consider adding additional elements to enhance the wind rose diagram, such as annotations indicating specific wind events or overlays of topographic features to show how terrain influences wind patterns.

By following these steps, you can create a comprehensive wind rose diagram to visualize and analyze wind patterns at a specific geographic location.

To construct this diagram, draw a small circle and write the number of calm days in it. Then draw straight lines in the given directions at the center of the circle. Choose an appropriate scale and draw each line from the circumference of the circle in that direction. Label the days of the wind. Name each direction and draw a scale below the diagram.

Example- Draw a wind diagram of Chennai with the help of the following data.

Table No. 3.9

Wind Direction	N	NE	E	SE	S	SW	W	NW	Calm
No. of Days	36	41	55	79	19	20	91	22	22

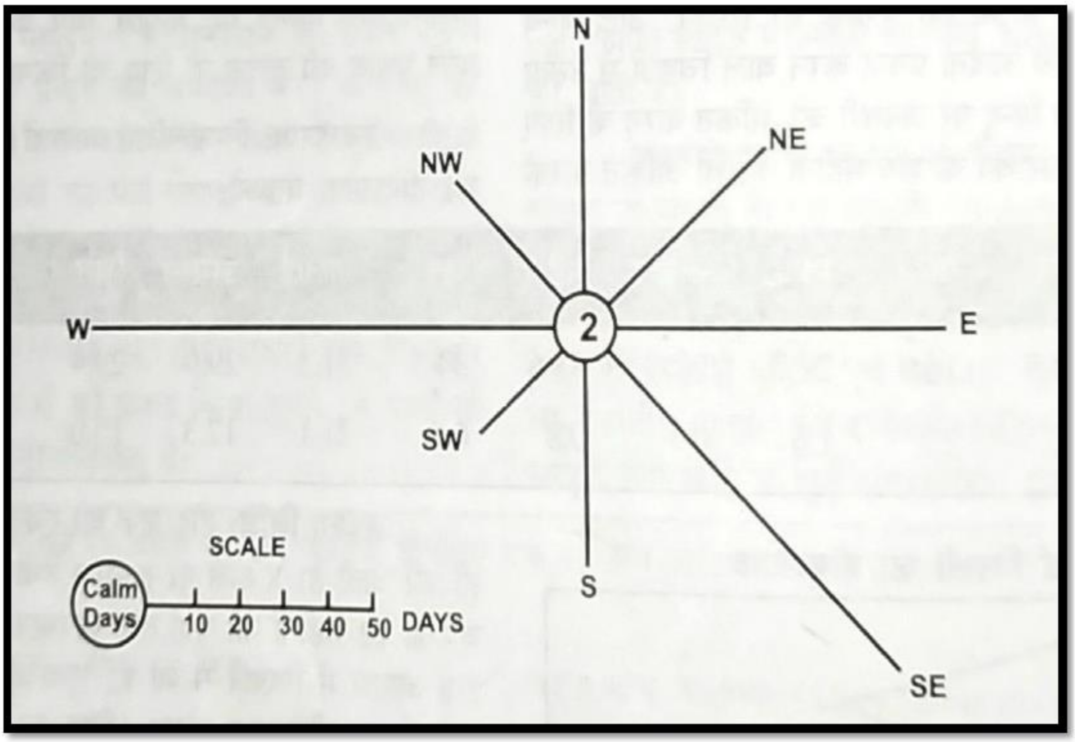


Fig 3.7: Wind Rose Diagram, Source: Google

Here's how a rose diagram typically works:

Circular Plot: The rose diagram is usually presented as a circular plot, with directional sectors representing different compass directions (e.g., north, northeast, east, etc.) evenly spaced around the circumference of the circle.

Frequency Distribution: The length of each segment or "petal" in the diagram represents the frequency or percentage of time that the wind or other phenomenon blows from that particular direction. The longer the petal, the more frequent the occurrence of winds from that direction.

Colour or Shading: Colours or shading may be used to enhance the visualization, with different colours representing different wind speed ranges or other relevant attributes.

Wind Speed Sectors: In addition to direction, some rose diagrams may also include sectors representing different wind speed ranges. This adds an additional dimension to the visualization, providing insights into the intensity of the winds from different directions.

Key or Legend: A key or legend is typically provided to explain the meaning of the different sectors, colours, or symbols used in the diagram.

Rose diagrams are valuable tools for understanding regional wind patterns, which can have significant implications for activities such as agriculture, construction, aviation, and maritime navigation. They allow meteorologists and other researchers to quickly identify prevailing wind directions, patterns of variability, and potential sources of local wind disturbances. Additionally, they can be used to compare wind patterns between different locations or over different time periods.

ii) Simple Star Diagram

Simple Star Diagram- As the name suggests, it is a simple diagram to show the direction of the winds. To construct it, join all the eight directions of the wind diagram with straight lines. In this way an octagonal shape will be obtained. It looks like a star, due to which it is called a simple star diagram.
Through this diagram, the wind that blows at a place throughout the year the number of direction and the number of calm air are revealed. i.e. air how many days from which direction did it move and for how many days was it calm? In this diagram calm wind is shown by making a small circle in the centre and according to the measurement of the direction of wind, it is shown in each direction. The octagonal diagram. It is called Star Diagram. In this diagram the eight directions of the wind North, North East, South, South West and North West are revealed. When needed all the eight directions like North-North-East, North-East etc. can also be shown.

Example- With the help of the following data makes a star-diagram of Delhi.

Table No. 3.10

Wind direction	N	NE	E	SE	S	SW	W	NW	Calm
No. of Days	33	22	24	29	17	16	102	101	21

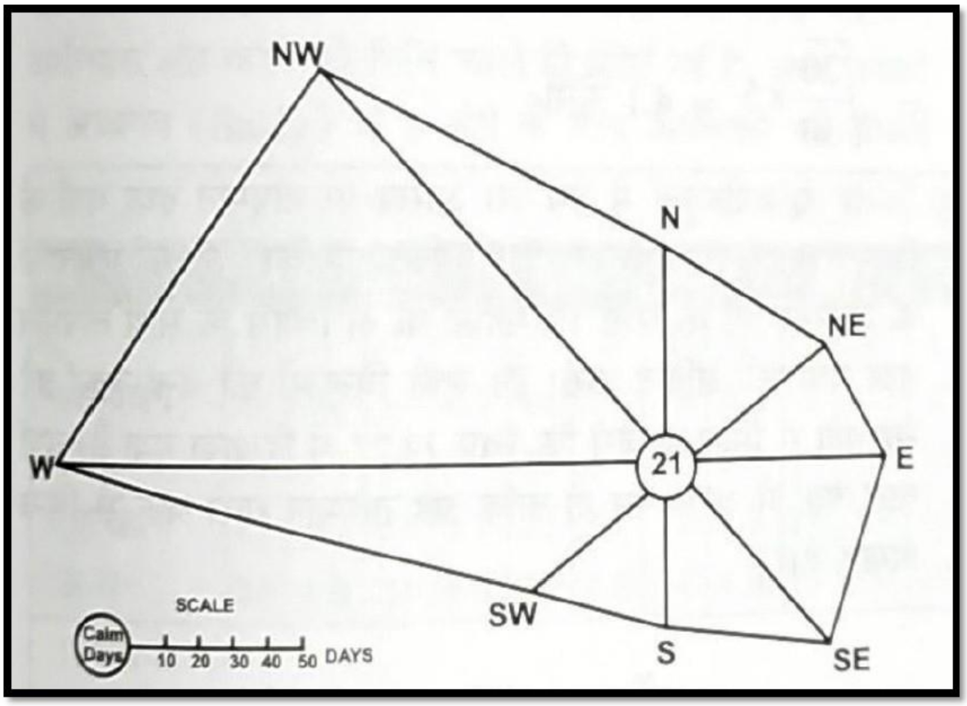


Fig3.8: Simple Star Diagram, Monthly wind direction and wind volume (as a percentage) are given for Jodhpur, Rajasthan throughout the year, draw a star diagram

Scatter Diagram

What is a Scatter Diagram?

- A scatter diagram, also known as a scatter plot, scatter graph and correlation chart is a graphical representation of data points in a Cartesian coordinate system.
- It helps visualize the relationship between two variables by plotting their corresponding values.
- The pattern of points on the scatter plot reveals the relationship between the variables.

Scatter Diagram Correlation Patterns

The degree to which the variables are related to each other depends on how the points are scattered over the chart. The more the points plotted are scattered over the chart, the lesser is the degree of correlation between the variables. The more the points plotted are closer to the line, the higher is the degree of correlation. The degree of correlation is denoted by "r".

Types of Scatter Diagram

i)Positive correlation

ii) Negative correlation

iii) No correlation

Positive correlation: The scatter plot with a positive correlation is also known as a "Scatter Diagram with Positive Slant." In this case, as the value of X increases, the value of Y will increase too, which means that the correlation between the two variables is positive. If you draw a straight line along the data points, the slope of the line will go up. For example, as the days planting are increases the heights of plants is also increases..

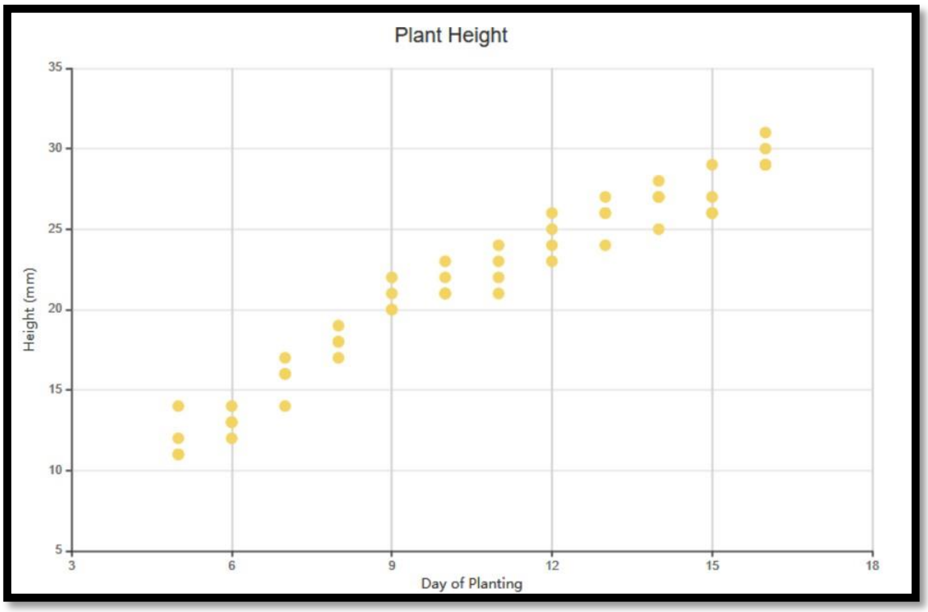


Fig 3.9: Positive Correlation, Source: Google

ii) Negative correlation: The scatter plot with a negative correlation is also known as a "Scatter Diagram with a Negative Slant." In this case, as the value of X increases, the value of Y will decrease. If you draw a straight line along the data points, the slope of the line will go down. For example, if the cycle time of a workflow goes up, the number of tasks completed will go down.

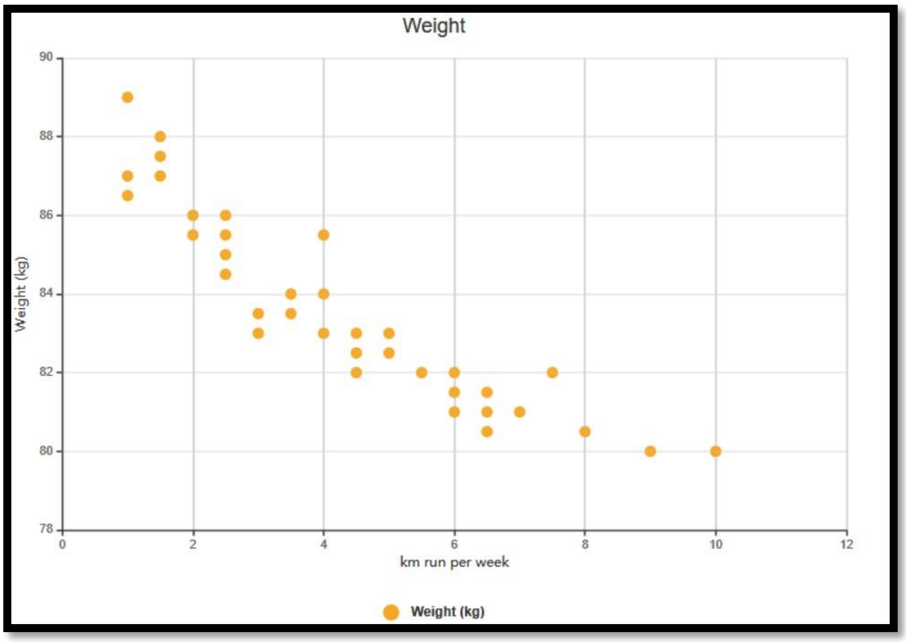


Fig 3.10: Negative Correlation, Source: Google

iii) No correlation: The scatter plot with no correlation is also known as the "Scatter Diagram with Zero Degree of Correlation." In this case, the data point spreads so randomly that you can't draw a line through the data points. You can conclude that these two variables have no correlation or zero degrees of correlation. For example, if the weather gets hotter, we can't conclude that the sales of wooden chairs will go up or down because there is no correlation between the two variables.

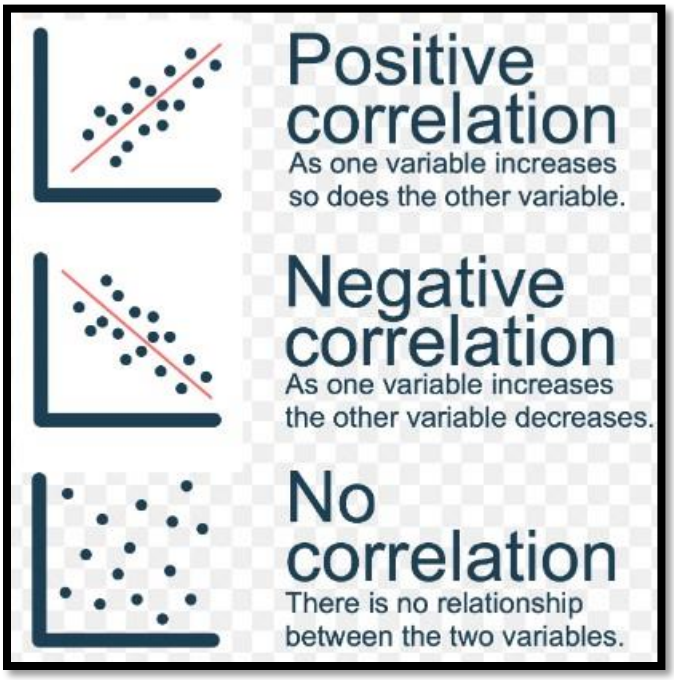


Fig 3.11: No Correlation, Source: Google

Creating a Scatter Diagram:

Here are the steps to create a scatter diagram:

a) Identify Variables:

- Determine which variables you want to compare. One variable will be the independent variable (usually plotted on the X-axis), and the other will be the dependent variable (usually plotted on the Y-axis).

b) Draw Axes:

- Set up the X-axis and Y-axis with appropriate scales.
- Label the axes clearly.

c)Plot Data Points:

- Plot each data point on the graph, representing the paired values of the two variables.
- Label each point to identify its significance.

d) Line of Best Fit:

- Draw a line of best fit through the data points.
- The line represents the general trend or average behaviour of the variables.
- The slope and direction of the line indicate the strength and direction of the correlation.

Interpretation: Analyze the scatter plot to determine the type of correlation:

- Positive Correlation: Points tend to form an upward-sloping line.

- Negative Correlation: Points tend to form a downward-sloping line.
- No Correlation: Points are scattered randomly without a clear pattern.

Example: Let’s consider an example related to rivers:

- Suppose we want to investigate the relationship between the width and depth of a river as it moves from its source to its mouth.
- We collect data on river width and depth at various locations along the river.
- After plotting the data points, we draw a line of best fit to understand the correlation.

Reading the Scatter Graph:

i)The line of best fit indicates the correlation strength between the two data sets.

ii)Positive correlation suggests that as river depth increases, so does its width.

iii) Scatter diagrams help geographers analyze relationships in various contexts, such as climate patterns, urban planning, and environmental studies.

Ergographs

Ergographs are graph that shows the relation between the season, climate and crops. The term ergo means work derived from the Greek word erogon. The name Ergograph was coined by Dr. Arthur Geddes of the University of Edinburgh. Ergograph is a graph with multiple variables showing economic activities performed in a particular season. The relation between human activities or climatic factors and a seasonal year can be shown graphically by Ergographs. It can be shown on the bar graph or line graph. In cartesian form, months are marked on the horizontal axis with the acreage of different crops, and the vertical axis shows the monthly average temperature, rainfall, and humidity. Temperature, precipitation, and humidity are chosen as they have a direct impact on crop production. The rectangular blocks show the net acreage of the crops, where the length of the block shows a growing season, and its breadth is calculated to a selected scale. In Ergograph, three elements are shown differently. First, the cultivated area will be demonstrated through the rectangular diagram. Second, rainfall can be shown through a bar graph, whereas line graphs can show temperature or humidity.

Construction:

Construct an Ergograph of Station A representing the data given below:

Table No. 3.11

Months	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature (degree Celsius)	13	15	22	27	30	31	30	29	28	27	18	15
Rainfall(mm)	29	28	9	11	14	60	299	298	290	15	7	9

Table No. 3.12

Crops	Growing	Net Sown Area (in '00 Hectare's)
Rice	May to October	28.4
Wheat	November to April	50.6
Cotton	June to December	13.4
Sugarcane	May to December	2.2

Net acreage of various crops along with growing season:

Steps:
1. On x-axis represent 12 months of the year
2. On the y-axis, the rainfall and temperature are represented.
3. A suitable scale is selected to represent these two factors. Here, the scale used for temperature is 1 cm = 5 degrees Celsius and for rainfall is 1 cm = 30 mm.
4. Temperature is shown through a line graph and rainfall through continuous bars.
5. The net sown area is shown through rectangles at the lower end of the graph. Again, a scale is fixed; here, the scale used is 1 cm = 200 hectares. The rectangle should be of the proportionate width and cover the months of the growing season, as shown in figure 11.

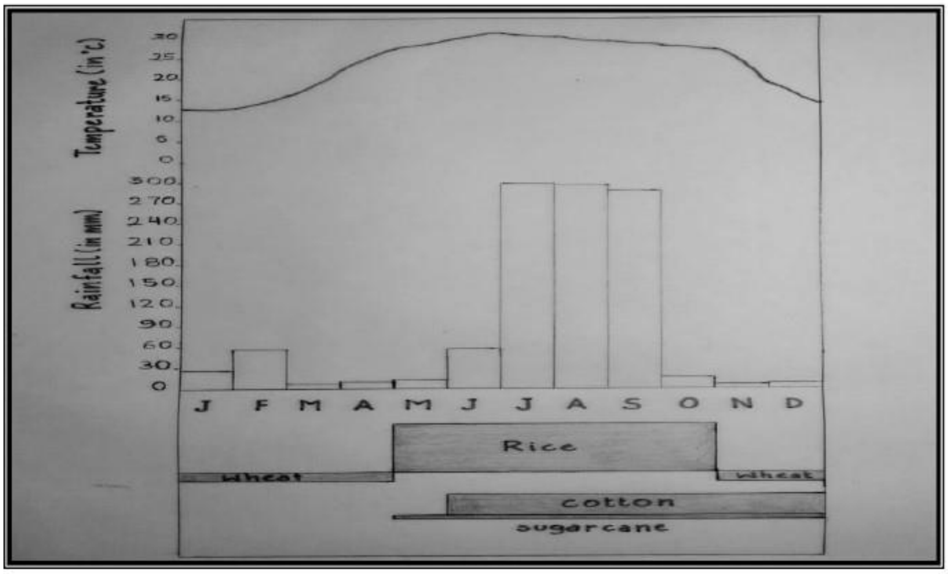


Fig 3.12: Ergograph

Application:

a) Ergographs are particularly useful for studying seasonal variations in human behavior, agricultural practices, and climate-related phenomena.

b) Researchers can analyze patterns such as changes in work hours, crop planting and harvesting, and other cyclic activities.

Significance:

a) Ergographs help us understand how human activities and natural processes align with the annual rhythm.

b) By visualizing these patterns, we gain insights into the impact of seasons on our lives and the environment.

Flow diagram map

The term 'Flow map' originally comes from cartography. It is a particular combination of maps and flow diagrams, where the width of the arrows is proportional to the flow rate.

Definition and Purpose:

A flow map visually illustrates the flow or movement of various entities, such as people, goods, animals, or information, across geographical regions.

A flow map is a specialized thematic map designed to depict the movement of quantities, such as information, weather events, people, animals, or physical objects, between different locations. It also indicates the volume of these movements

These maps help us understand patterns of connectivity, migration, and distribution by emphasizing the direction and magnitude of movement.

Design Elements:

(i) Linear Symbols: Flow maps use linear symbols (usually lines or arrows) to depict the movement between locations.

(ii) Width Proportional to Flow: The width of these lines represents the quantity of flow. Wider lines indicate higher volumes of movement.

(iii) Hybrid Nature: Flow maps are a hybrid of traditional maps and flow diagrams, combining spatial information with flow visualization.

Types of Flow Maps:

(i)Origin-Destination Maps: Show the source and destination of movement (e.g., migration patterns, trade routes).

Fig 3.13: Origin-Destination Map, Source: Google

(ii)Route Maps: Display the path along which movement occurs (e.g., transportation routes, migration corridors).

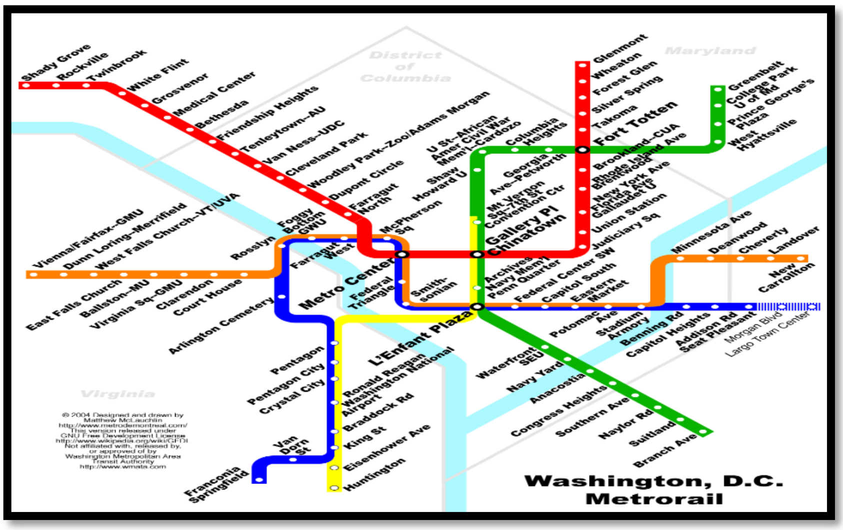


Fig 3.14: Route Map, Source: Google

The amount of migration in a particular flow line is represented by its size: the more thick the line is, the stronger the migration wave it stands for. Flow lines are drawn from a point of origin and branch out with arrows used to show the direction of movements.

(iii) Continuous/Mass Flow Maps: Represent continuous flows (e.g., river currents, air masses).

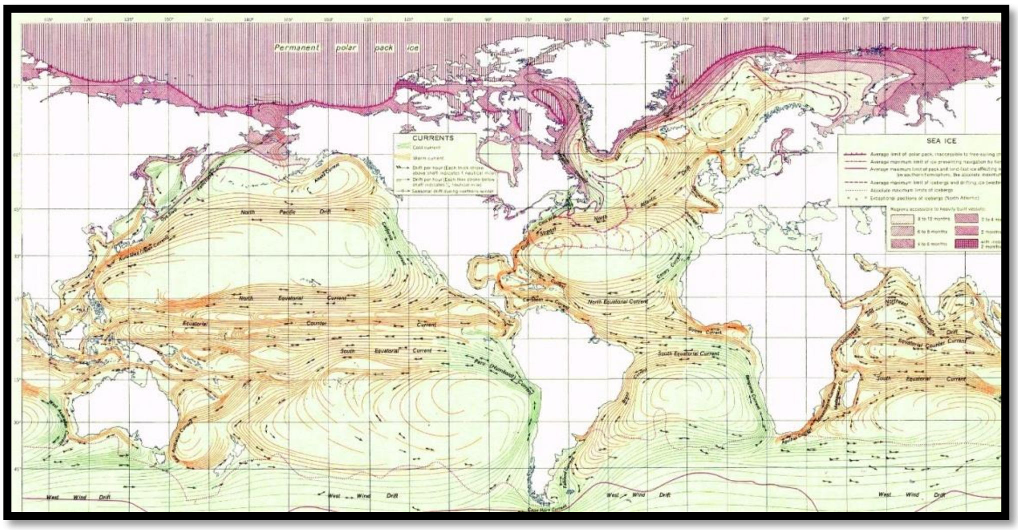


Fig 3.15 Continuous Mass Flow Map, Source: Google

(iv) Weight Scaling: Adjust line width based on the magnitude of flow (e.g., thicker lines for major highways).

Historical Context: The earliest known flow maps date back to the 1830s. Engineer Henry Drury Harness created maps showing cargo traffic by road and canal in Ireland.

Charles Joseph Minard, a renowned cartographer, perfected flow maps during the 1850s and 1860s. His map of the French invasion of Russia in 1812-1813 is considered a masterpiece..

3.4 SUMMARY

Combining correlation analysis with various cartographic representations enhances the understanding of socio-economic data. Correlation analysis identifies relationships between variables, while cartographic methods visually represent these relationships and distributions, making complex data more accessible and interpretable. For example, Pearson's or Spearman's correlation results can be effectively displayed on scatter diagrams, and socio-economic indicators can be vividly illustrated using choropleth maps or flow line diagrams. This integrative approach is crucial for fields such as geography, urban planning, and public policy, where spatial and statistical analysis intersect.

3.5 GLOSSARY

- **Karl Pearson's Correlation Coefficient (r):** A measure that quantifies the strength and direction of a linear relationship between two continuous variables.
- **Spearman's Rank Correlation Coefficient (ρ or r_S):** A non-parametric measure of rank correlation that assesses how well the relationship between two variables can be described using a monotonic function.

- **Flow Line Diagrams:** Diagrams that use lines of varying thickness to represent the flow of goods, people, information, or other entities between locations.
- **Ergographs:** Graphs that show the variation of a variable over time, often in the form of a bar or line chart.
- **Star Diagrams:** Diagrams that use lines radiating from a central point to represent multiple variables, with the length of each line proportional to the variable's value.
- **Choropleth Maps:** Maps that use colour shading within predefined areas (such as countries, states, or regions) to represent different data values or classes.
- **Proportional Squares:** Squares of varying sizes representing data quantities, where the area of the square is proportional to the value it represents.
- **Proportional Circles:** Circles of varying sizes representing data quantities, where the area of the circle is proportional to the value it represents.
- **Dot Maps:** Maps that use dots to represent the presence, quantity, or value of a phenomenon, with each dot representing a specific number of occurrences.

3.6 ANSWERS TO CHECK YOUR PROGRESS

1.Do you that Karl Pearson's Correlation Coefficient is used in various fields to understand the degree of association between variables, such as in, geography, economics, biology, and social sciences.

2.Do you that the coefficient (r) ranges from -1 to +1, where +1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

3.7 REFERENCES

- Tideman, E. M. (1996). Watershed Management Guidelines for Indian Conditions. Omega Scientific Publishers, New Delhi.
- Mishra, R.C, Practical Geography, Vishwabharti Publication, 4378/4b Ansari Road, Dreyaganj New Delhi-110002
- Shukla, S.M, & Sahai, S.P, Business Statistics, Sahitya Publications, Hospital Road, Agra-282003.
- https://www.simplilearn.com/tutorials/statistics-tutorial/spearmans-rank-correlation
- https://www.slideshare.net/slideshow/pearsons-correlation/250692763
- https://www.youtube.com/c/esriindia
- https://spatialthoughts.com/2024/05/13/30daysofqgis/
- https://www.simplilearn.com/tutorials/statistics-tutorial/spearmans-rank-correlation
- https://www.brainkart.com/article/Dot-Method_41157/
- https://www.youtube.com/watch?v=P53ox8B73ZU
- https://online.visual-paradigm.com/knowledge/data-visualization/what-is-scatter-diagram/
- https://www.internetgeography.net/scatter-graphs-in-geography/#:~:text=A%20scatter%20graph%20is%20used,the%20points%20have%20been%20plotted.
- https://www.slideshare.net/slideshow/pearsons-correlation/250692763
- https://www.brainkart.com/article/Dot-Method_41157/

- https://www.youtube.com/watch?v=P53ox8B73ZU
- https://en.wikipedia.org/wiki/Flow_map#:~:text=of%20several%20types.-,Origin%2Ddestination%20map,often%20straight%20or%20slightly%20curved.

3.8 TERMINAL QUESTIONS

Long Questions

1.What is Karl Pearson's Correlation Coefficient and Spearman's Rank Correlation Coefficient? explain them in detail.

2. Explain the given cartography & representation of socio-economic data in detail: dots, proportional squares, circles, spheres, choropleth, star & scatter diagrams, ergographs, flow line diagrams.

3. Given the formula for Karl Pearson's correlation coefficient r:

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

where n is the number of pairs of scores, x and y are the individual scores, $\sum xy$ is the sum of the products of paired scores, $\sum x$ and $\sum y$ are the sums of the x scores and y scores respectively, $\sum x^2$ and $\sum y^2$ are the sums of the squared xx scores and yy scores respectively, calculate the correlation coefficient given the following data set:

X	2	4	6	8
Y	3	5	7	9

Show all steps in your calculation.

Short Questions

1 What is Karl Pearson's Correlation Coefficient?

2.What is Spearman's Rank Correlation Coefficient?

3.How does Spearman's rank correlation coefficient differ from Pearson's correlation coefficient?

4. What is a dot map, and when is it typically used?

5. In what scenarios would you prefer using Spearman's rank correlation over Pearson's correlation?

6. How do proportional squares and proportional circles differ in representing data quantities?

7. What type of data visualization is ideal for showing variations in socio-economic indicators like income levels or election results within predefined areas?

8.What is the primary advantage of using a scatter diagram in data analysis?

9.What is Ergograph and what kind of data is best visualized using an ergograph?

10.For what purpose are flow line diagrams most commonly used?

11. What is scatter diagram and what type of diagram uses lines radiating from a central point to compare multiple variables simultaneously?

Multiple choice Questions

1. What does Karl Pearson's correlation coefficient measure?

A) The difference between two variables

B) The strength and direction of a linear relationship between two variables

C) The average of two variables

D) The rank order of two variables

2. What is the range of Pearson's correlation coefficient?

A) -1 to 1

B) 0 to 1

C) -100 to 100

D) 0 to 100

3. Which correlation coefficient is used for ranked data?

A) Pearson's

B) Spearman's

C) Standard deviation

D) Mean

4. Spearman's correlation coefficient assesses relationships based on what type of data?

A) Interval data

B) Nominal data

C) Ranked data

D) Ratio data

5. Spearman's rank correlation coefficient of 1 indicates what kind of relationship?

A) No correlation

B) Perfect positive rank correlation

C) Perfect negative rank correlation

D) Moderate correlation

6. In a proportional circle map, what does the size of the circle represent?

A) The geographical area

B) The population density

C) The data value it represents

D) The distance between data points

7. What type of map uses colour shading to represent data values?

A) Scatter diagram

B) Dot map

C) Flow line diagram

D) Choropleth map

8. Which diagram uses lines radiating from a central point to compare multiple variables?

A) Star diagram

B) Scatter diagram

C) Ergograph

D) Flow line diagram

9. Which type of diagram is commonly used to show variations over time?

A) Star diagram

B) Ergograph

C) Dot map

D) Proportional square

10. What do flow line diagrams represent?

A) The correlation between two variables

B) The rank order of variables

C) The distribution of population

D) The flow of goods, people, or information

11. What type of cartographic representation would be best to show trade routes?

A) Choropleth map

B) Dot map

C) Flow line diagram

D) Scatter diagram

12.Which map type is ideal for showing population density?

A) Flow line diagram

B) Dot map

C) Scatter diagram

D) Star diagram

13. Which correlation measure would you use if your data are not normally distributed?

A) Pearson's

B) Spearman's

C) Mean

D) Standard deviation

14. If Pearson's correlation coefficient is 0, what does it indicate?

A) No linear relationship

B) Perfect positive correlation

C) Perfect negative correlation

D) Strong correlation

15.What is the primary purpose of a choropleth map?

A) To show the rank order of data

B) To represent data values using colour shading

C) To display individual data points

D) To illustrate the flow of goods and services

Answers) 1.B, 2. A, 3.B,4.C,5. B, 6.C, 7.D,8. A, 9.B, 10. D, 11.C, 12.B, 13.B, 14.A, 15.B

UNIT-4 QUANTITIVE TECHNIQUES IN GEOGRAPHY, CONTOURING AND DETERMINATION OF HEIGHT WITH INDIAN PATTERN CLINOMETERS

4.1OBJECTIVES

4.2 INTRODUCTION

4.3. CONTOURING AND DETERMINATION OF HEIGHT

4.4 SUMMARY

4.5 GLOSSARY

4.6 ANSWER TO CHECK YOUR PROGRESS

4.7 REFERENCES

4.8 TERMINAL QUESTIONS

4.1 OBJECTIVES

After studying this chapter you will be able to:

- Understandings the construction process of the contour line
- Explain the utility of the contour line.
- To understand the method of use of the Indian Clinometers' Survey instrument

4.2 INTRODUCTION

A contour line is an imaginary line joining land areas of equal height determined by a survey from sea level. Which is simple and useful in understanding the physical structure of a particular place on the topographic map of the earth's surface? Contour lines have been very useful from ancient times to the present, but with the inclusion of many new techniques in the technical field, photography is now being used. Contour lines are being determined very easily through northern photography and remote sensing techniques. The unit of measurement is the metre, which is drawn at vertical intervals of 20, 50, and 100 metres in a contour map. Where there is a difference in the horizontal distance between two contour lines according to the conditions of the slope of a particular area.

Where the vertical distance between them remains unchanged. The horizontal distance is known as the horizontal equivalent (HE). Whatever is low relative to the slope continues to increase. Extreme caution has to be taken in determining the contour lines because, due to being thin, they get mixed together, due to which the determination of relief cannot be shown clearly and the actual shape of the landform is not revealed accurately. Thus, contour lines are known as isohypses. This is determined by meeting all the points of equal height or depth above the standard base level or international sea standard pools determined by the Survey Department of the Government of India. Apart from this, there is no other basis for their determination.

Generally, maps contain only spatial heights, and marking is done through contour lines only. Contour lines are created only with the help of spatial heights. Till two-three decades ago, heights obtained from survey instruments like Indian clinometers, dumpy level, sextant, and theodolite were used to draw contour lines. But with modern technical knowledge, it is now being prepared in less time and at a lower cost than computer models. Contour lines are used to reveal various relief forms.

Level measurement survey instrument: The Indian clinometers have been the simplest instrument in determining spatial heights as compared to all level measurement survey instruments; it is light and easy to carry from one place to another. Easily carried by a surveyor to any inaccessible place, this instrument was manufactured by the Survey of India, from which it was named Indian Clinometers'. This instrument has been extremely useful for measuring small heights and depths, from which spatial heights and depths can be read from the values of angles and natural tangent values. The highest value in degrees is 22, and the natural tangent value is 0.4. This instrument is made of iron metal, which is 22 cm long and 2 cm wide.

4.3. CONTOURING AND DETERMINATION OF HEIGHT

4.3.1- Contouring- ***Meaning*** - A contour line is made up of two words: even and high lines or contours that represent any physical quantity. Reveals similarity in (altitude, temperature). For example, contour lines on a map that join land areas of equal height are imaginary lines that reveal the spatial height of physical relief landforms from a standard base plane, whose horizontal gradient varies depending on the slope of the surface. On the basis of which the physical form can be easily identified.

In definition - "The imaginary lines meeting parts of equal height above sea level are called contour lines." Thus, the contour line is a graphical display of places of equal height on the surface, which is drawn with reference to a base plane selected from the sea level.

4.3.2 Contour lines

Generally, a line intersected by a plane on the surface or paper is known as a contour line. Contour lines show the height and depth of all the relief areas in the first, second, and third orders of the surface. To make a contour map of any part of the earth's surface, we should mainly keep two things in mind: (i) as the first task, the height of different places in the survey area should be determined from the standard base level. (ii) A rough plan of the terrain determining the height of the contour lines should be prepared.

In this way, with the help of contour lines, the relief of the land surface is displayed on the map. Which all the time In science, there is a standard method of relief representation that provides the basis for many relief representations. These are obtained by interpolating local heights. Through contour lines, the map shows the parts of the land of the entire world with equal height above sea level, which is measured from a fixed imaginary point. The difference between two contour lines in a contour map is called the contour line interval.

But sometimes this interval is increased by less than 20, 50, or 100 metres in the maps of mountainous areas made on a small scale, whereas in maps made on a larger scale, the contour line interval is done as per national and international standards. The number of contour lines determines the gradient. Where the number of contour lines is greater and they are close together, there is a steep slope, while on the other hand, if there is more distance between the contour lines, then it is called a medium or low slope. Let's reveal. This gradient determines the conditions for terraced farming in mountainous areas; contour farming has special importance in weakening the erosion forces. Weather forecasting and agricultural patterns in different areas are also done on the basis of the height of the contour lines, which makes it easier for humans to carry out similar agricultural activities in different areas with the same height because the contour lines are on the ground. The slopes and landforms are displayed. Which is explained by figure number 6.1?

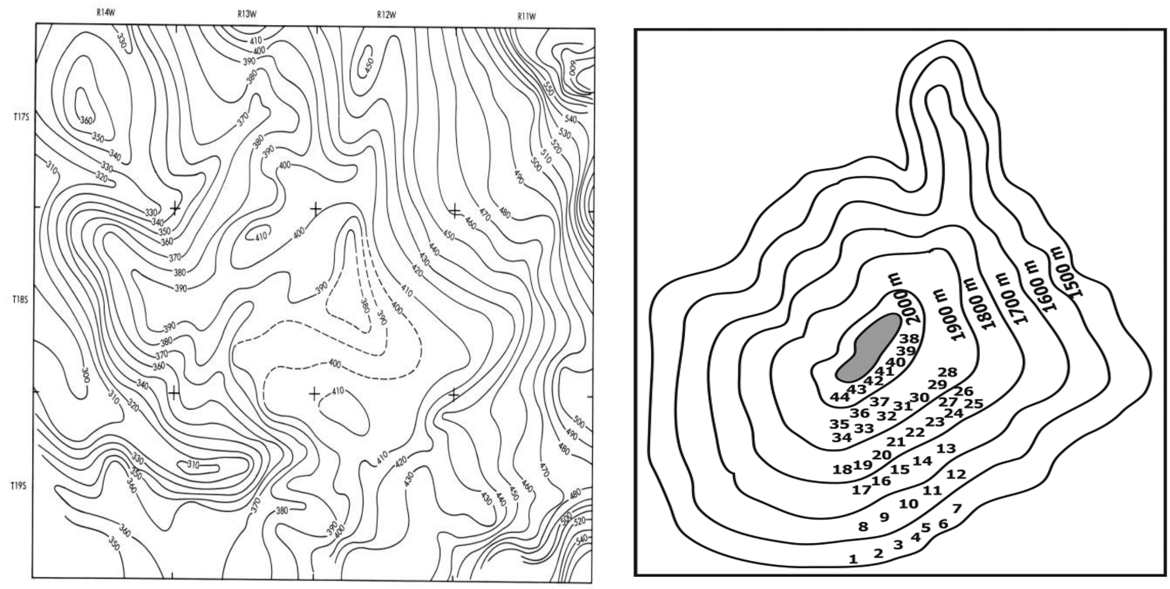


Fig. No. 4.1 Source: Google

4.3.4 Precautions in determining contour lines

While constructing a linear map, a surveyor should construct it carefully and with concentration, because a small mistake by the surveyor will not reveal the actual points of many places on the ground surface. Due to this, wrong results persist for a long time, even in the formulation of far-reaching policies. Mainly, the surveyor and cartographer should take the following precautions while determining the contour lines:

1. The surveyor should already know the height above sea level of the land being surveyed.
2. Before constructing the contour line, the preparation of the final contour map should be started only after self-observing the entire area and preparing a rough plan.
3. Contour lines should not have straight lines, sharp modes, or continuous small modes.
4. In contour maps with long and high-resolution features like the Himalayan Andes, the contour interval has to be kept 50–100 metres longer.
5. The scale line interval of the contour map has to be written clearly and according to the rules of cartography.
6. One should always try to keep the shape of the contour line small or circular.
7. If there is a deep place in determining the social lines, then shading (small lines) should be done at that place.
8. One should try to write as much as possible of human social lines (according to the space of paper).

4.3.4- Characteristics of contour lines

Contour lines have their own special identity, the main characteristics of which are explained in the following points.

1. Contour lines connect the land areas with the same height above sea level.
2. Contour lines and their shapes indicate the slope, height, and depth of the land structure.
3. Contour lines always remain neutral; they never intersect each other.
4. When two or more contour lines meet each other, vertical shapes such as waterfalls and curves are displayed.
5. Dense contour lines. Sharp pulses and rare, i.e., less watery contour lines, indicate gentle slopes.
6. Contour lines are most useful in representing the relief of a slope.

7. The contour line forms a right angle towards the direction of dip on the earth's surface and at the surface (in the direction of water currents).
8. The contour line is not able to make a real and very weak branch everywhere.
9. Contour Lines: In the places where they cross rivers and water streams, they are inclined upwards.

4.3.5 Process of determining height by contour lines-

Contour lines are lines drawn according to a set standard. In which the spatial height is written in bold and large size. For the convenience of calculation, every fifth point in the map is made thick. The height is calculated on the basis of the interval values of two similar lines. Suppose we find the altitude at places A, B, and C, which are situated at altitudes of 1000, 1500, and 2000 metres, respectively, and the contour line interval is 100 metres, which is a 1500-1000 = 500-metre difference between points A and B, which are 2000-1500 = 500 metres, whereas in points A and B there is a clash of 2000-1000 = 1000 meters.

If the height of different places has to be found, then it is very important to know the common longitude, such as to find the height of A.B.C. contour lines that are situated in different places and whose contour line interval is 100 meters.

4.3.6- Interpolation of contour lines-

The art of representing contour lines with the help of spatial heights determined by sea level is called contour line interpretation. Spatial heights are marked on the map with the help of a plane table. The more points of spatial height are taken on the plane, the greater the accuracy in drawing the contour lines in the map. And the contour line is drawn equally accurately. The contour line interval should be a fixed value (round figures) of 10, 20, 30, 50, and 100 meters. The interval of the contour line is determined by the interval of the lowest and highest height of the contour line. Supposing A and B are two contour lines with heights of 500 and 1000 metres, then the contour line interval is 1000 minus 500 = 500 metres, i.e., A contour line B contour is 500 metres below the line. As a general rule, there should be similarities between two contour lines.

4.3.7 Relief representation by contour lines

Till now, we have understood the construction, determination, and characteristics of contour lines well, but the details of how contour lines are used to determine the relief features of the first, second, and third types of relief forms It is determined by different topographies and types of slopes. Which are gentle slope, steep slope, concave and convex topography (cone shape, plateau, V-shaped valley, U-shaped valley, gorge, mountain spur, and waterfall), and other types of topography whose details are given in the above topography. While representing relief features with the help of contour lines, a cartographer has to especially keep in mind the actual shape of the ground and the height above sea level.

1. **Uniform slope:** when the angle of slope of the topography is very small and the distance between the contour lines is very large.
2. **Steep slope:** When the angle of slope of topography is high, the interval of contour lines between them is very small (close together). The slope is around 75°.
3. **Concave slope:** if the lower part of the topography has a gentle slope and the upper part has a big slope, then it is called a concave slope. The contour lines are far apart in the lower part and close together in the upper part.
4. **Convex slope:** The topography having characteristics opposite to the concave slope is called the convex slope. The contour lines are close together in the lower part and far apart in the upper part. Which is explained by figure number 6.2?

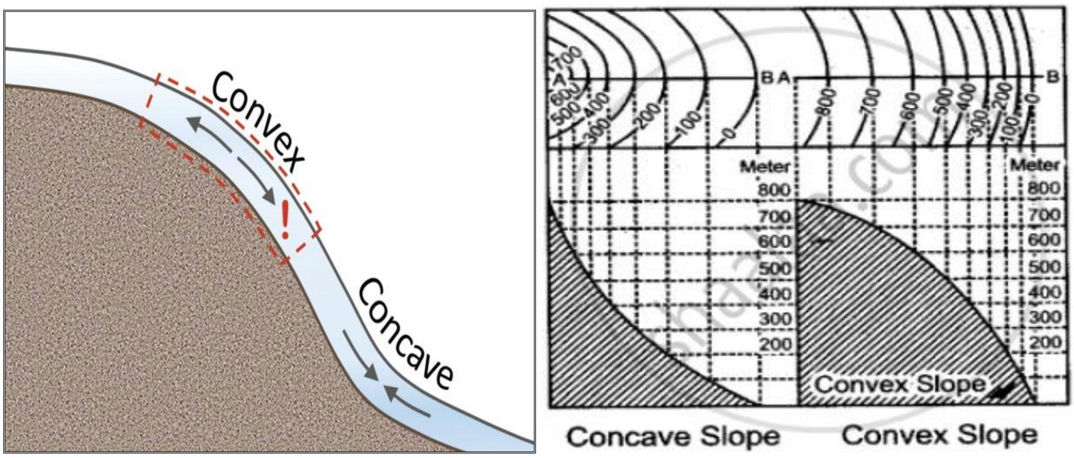


Fig No. 4.2 Source: Google

5. **Undulating slope:** The distance between the slope contour lines is also the contour lines are far apart at some places and close together at others.
6. **Terraced slope:** In this type of topography, stairs are formed at small distances. The contour lines are in the form of pairs, with approximately the same distance between the pairs. Which is explained by figure number 6.3?

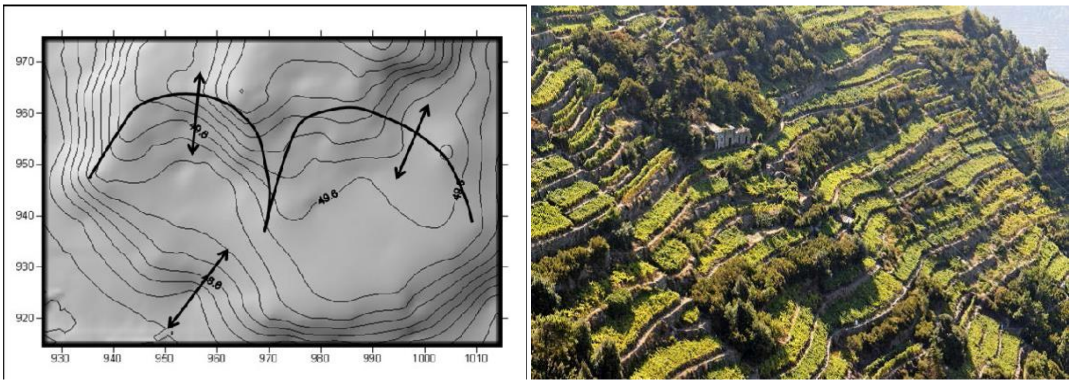


Fig. No. 4.3 Source: Google

7. **Conical hill:** A conical hill with a height of up to 1000 metres above sea level is called a conical hill. Contour lines are displayed at equal distances.
8. **Plateau:** The elevated land area above the adjacent areas whose peak is flat is called a plateau. In the plateau part, the contour line is close to the lateral edges and is flat in the central part. Which is explained by figure number 6.4?

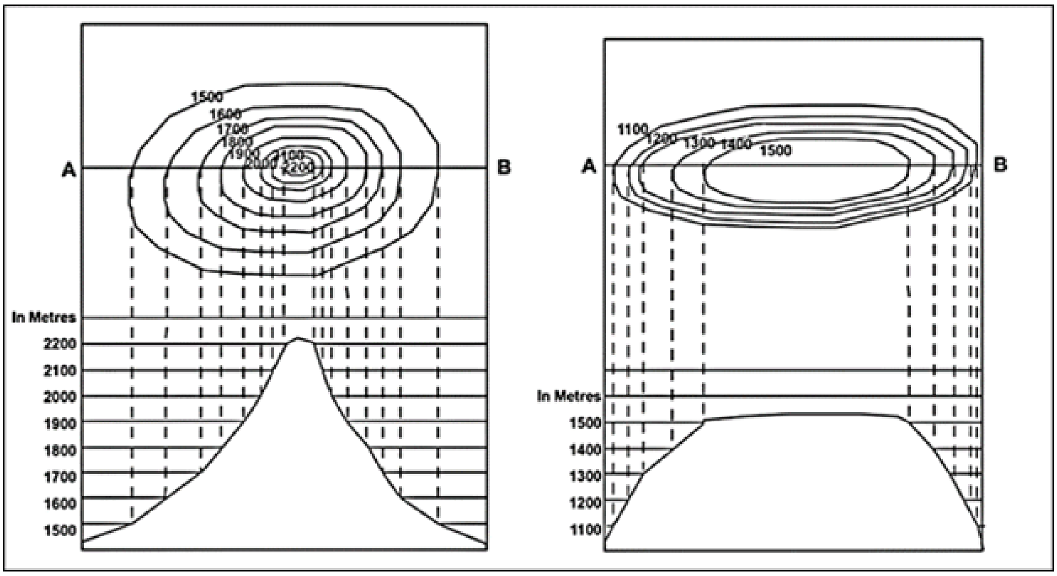


Fig. No. 4.4 Source- Google

9. **Ridge:** which the contour lines are downward in the central part and far apart in the adjacent flat area.

10. **Plain with Knoll:** A round-shaped hill of low height in the plains is called a knoll. Who it looks like dunes; the contour lines are far apart and circular in the middle.
11. **Escarpment:** A steeply sloping edge like a hill is called a ledge, in which the contour lines are made close to each other.
12. **V shaped Valley:** A long valley located between two mountains, in the middle of which a river flows, which looks like the English letter V, in which the contour lines get closer in the middle and farther apart on the sides. Which is explained by figure number 6.5?

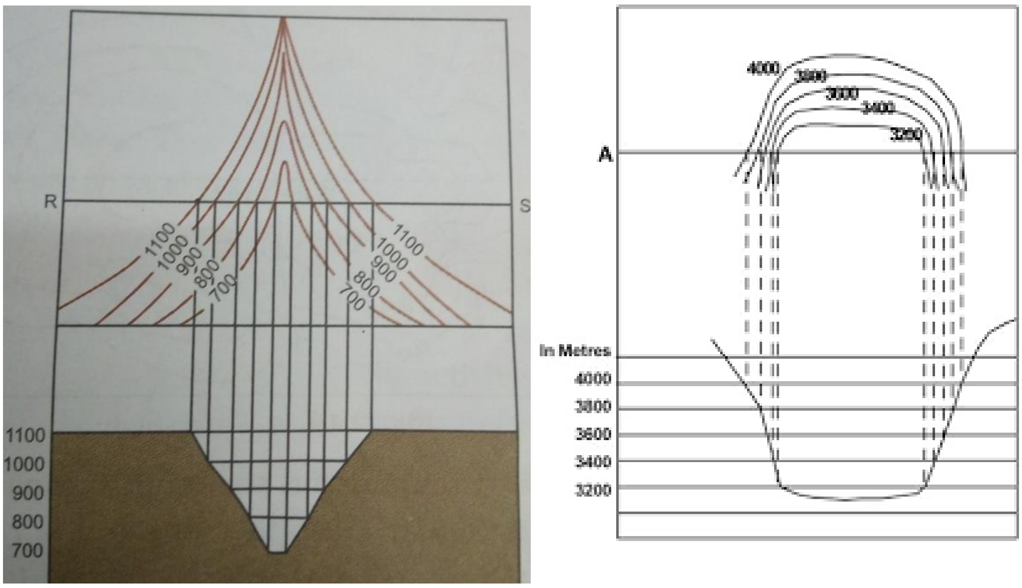


Fig. No. 4.5 Source: Google

13. **U-shaped valley: a** physical shape consisting of the English letter U formed between two mountains due to erosion by the glacial river in snow-poor areas, whose bottom is wide and the sides are steep. In the central part, the contour lines are close to each other at equal distances in the other parts.

Water fall: When waterfalls fall vertically from a sufficient height in a river, it is called water fall. In a waterfall, the contour lines meet each other. Which is explained by figure number 6.5?

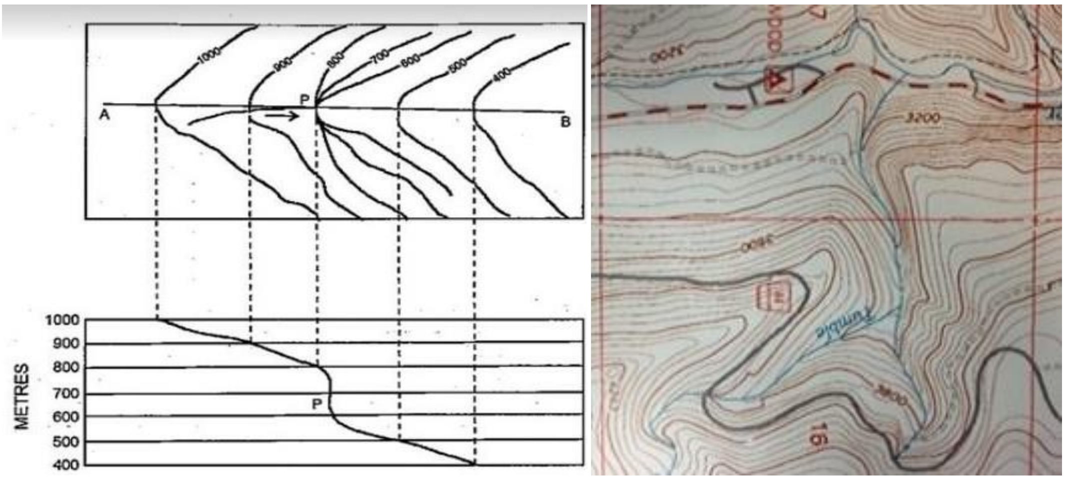
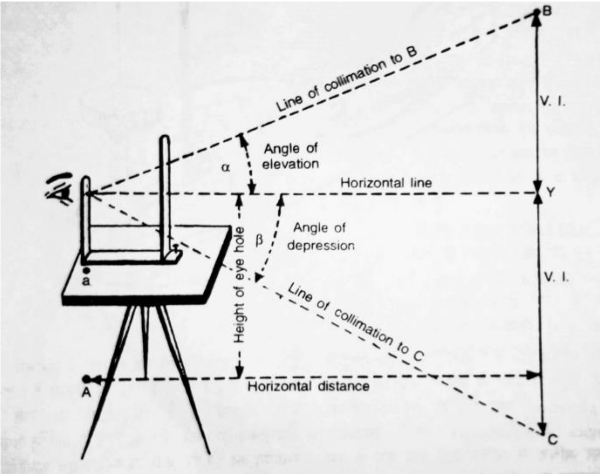


Fig. No. 4.6 Source: Google

14. **Gorge:** A deep, narrow valley with steep sides of a river is called a gorge. Here, the contour lines are very close.
15. **Lake:** A closed circular lake or trough is shown by contour lines. Where the distance of the contour lines increases from the centre outwards.
16. **Fiord Coast:** As a result of glacial action, the area with poor hem in the high latitude area is called Fiord, where the contour lines appear like a fan.
17. **Riya Coast:** The submerged part of the sea that goes below sea level due to erosion of the big river valley is called Riya Coast, and the society lines look like mountain ranges.

4.3.8 India pattern clinometers

Indian pattern clinometers were invented by the Indian Survey Department; hence, they were named Indian clinometers'. It is also known as a tangent clinometer. This instrument is a very useful and accurate instrument for measuring the height and depth of remote (not easily accessible) points. The Indian clinometer is used with the help of a plane table. This measures the angles of depression and elevation. This instrument is a modified form of the alidade, which is made from a heavy plate of simple brass metal. In the initial stages, the most expensive item in this survey instrument was the ivory button, which was considered an expensive instrument, but with changing circumstances and technological development, it is currently available at affordable prices. The main parts of the Indian Pattern Clinometers are as follows: Are described. Which is explained by figure number 6.6?



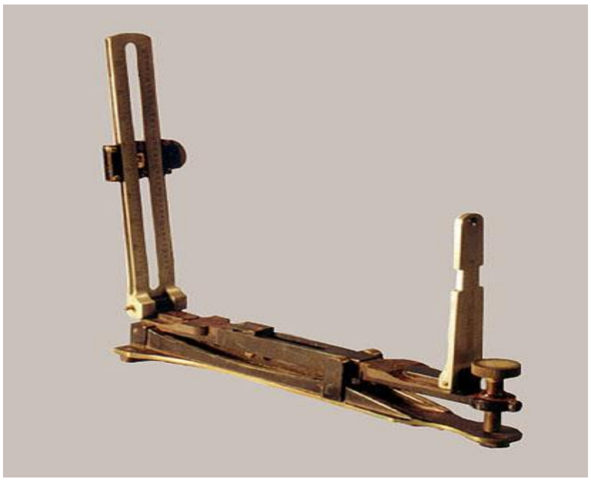


Fig. No. 4.7 Source: Google

1. **Base Plate:** This is a metal plate about 22 cm long and 2 cm wide at the bottom of the clinometers that is located on three buttons.
2. **Levelling screw strut:** The levelling of the Indian pattern clinometers placed in the plane table is done by the levelling punch; the work of stopping both ends of the base plate is done by the same strut.
3. **Brass bar:** There is a brass rod above the base plate on which the weight of the entire clinometer depends.
4. **Eye vane:** At the ends of the brass rod, there are two shaved vanes about 20 cm apart from each other, and the eye vane is the observation area with the help of which the target points are seen during the survey.
5. **Sliding Frame:** By sliding the frame itself, a transmitted point is established at the same position, which helps in reading the determined depression and elevation angles and natural tangents by targeting all the objects of the survey through the eye vane and object vane observation holes.
6. **Objects vane**: is a vertical strip with a slit cut in the middle. On its left edge, degrees are written, and on the right edge, signs of natural tangent are written, in the middle of which zero is written. Angles of depression are read from points at altitudes above this zero, and angles of depression are read from values below it.
7. **Other supporting equipment**: To conduct a survey using Indian pattern clinometers, a surveyor needs a plane table, tape measure, survey arrow, compass,

plumb tongs, simple rod, survey chart, board pin, pen, pencil, rubber, scale, and practice. A booklet is required.

4.3.9 - Features of Indian Pattern Clinometers-

The main features of Indian pattern clinometers are as follows:

1. It is a light, small, medium-sized survey instrument that can be carried from one place to another.
2. With this survey equipment, maps of even hard-to-reach points (out of reach) can be easily made.
3. In clinometers, angles up to 0.005 natural tangent value and 20 minutes can be read. With this instrument, maximum elevation angles of 22° can be measured.
4. Since the survey process of the Clinometers instrument is very simple, even a general geography student can make a map of the survey area with its help.
5. This instrument is the only excellent discovery of the Indian Survey Department, which developed the art of making maps even with limited resources in the field of survey.

4.3.10 Shortcomings of Indian Pattern Clinometers-

The biggest drawback of the Indian Pattern Clinometers is that they cannot measure angles above 22 degrees; hence, it is not possible to measure angles of elevation and depression higher than this.

4.3.11- Method of use of Indian Clinometers-

As a result of this being small field survey equipment, it is easy to use in surveys. First of all, the Indian Pattern clinometer instrument and its other supporting instruments (which are mentioned above) are selected in the selected survey area. After selecting a suitable place, the phone table is established by facing the instrument towards the south. After that, after selecting the correct levelling and direction, a survey instrument station is established, and after preparing a rough plan in the field book, the sent points determined with the help of clinometers are written in the form of A, B, and C lines, respectively.

Table No. 4.1 Field Book

Station	Angle of elevation or depression	Value of tangent	Height of eye hole from ground(HE)	Horizontal distance	Height of sea level
A	20°	0.4	1 M	150 M	200 M

Now, set the plane table level at station A and write it on the paper from point A. Enlarge both the lids of the clinometers, keep the eye vane facing you with the help of the labelling screw, and accurately level the clinometers with the bubble of the level table. Slide the sliding frame fitted in the second panel up and down with the help of a rock and pinion so much that the object becomes clearly visible through the eye vane, and if the horizontal wire is installed in the middle, then the same value will reveal the value of angle and natural tangents. Now they measured the height of the eye vane and the distance of the objects from the ground instrument station with a measuring tape. After that, the field book should be filled out, and a plan (map) of the area should be prepared by keeping the values obtained through the survey in the following formula:

Formula: vertical interval VI = horizontal distance (H.E.) × tan α + height of table.

By keeping the values of the field book in the above formula, the height of the clinometers (eye vane) is added, which will determine the height of the target object. If the depth of any place (angle of depression) has to be determined, then the eye vane table height will be reduced, and a map will be prepared after preparing a plan of the determined surveyed area.

4.3.12. Determination of height of points with clinometers-

When any target point on the ground is situated at a great distance from the instrument station or it is not easy to reach that point, it is difficult to measure its horizontal distance. In such a situation, the height and altitude of the surveyed area can be determined by adopting the following procedure for Indian Pattern Clinometers: Depth can be determined. The method of using Indian Pattern Clinometers has made it clear that under this process, the instrument is first installed at a designated place, and two stations (A and B) are selected, which will determine the altitude of points a and b. Now install the plane table at A station, find the vertical angle A at B point, and find the height of the instrument. Next, select the C station; keep it parallel to A at the same height, and measure the distance between A and C. at C station again. The values of vertical angles are also determined. By filling in the values (A and C stations) in the exercise book, the height of the point is also determined with the help of the following formula: Which is explained by figure number 6.7?

Table No. 4.2 Field Book

Station	AC Horizontal distance	Angle of elevation or depression	Value of tangent	Height of eye hole from ground(HE)	Height of sea level
A	66 M	$20^{\square}$	0.4	1 M	200 M
B	-	25^{0}	0.3	1 M	200 M

$$VI = \frac{\text{AC Distance} \times \tan\alpha \times \tan\beta}{\tan\beta - \tan\alpha} + - \text{Height of table}$$

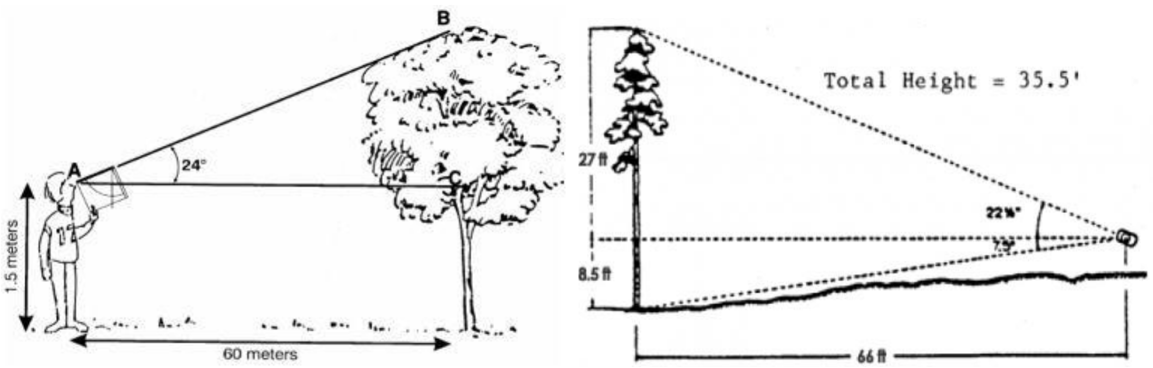


Fig. No. 4.8 Source: Google

By placing the values of all the surveyed points in the above formula, a surveyor gets the height of the transmitted point of the surveyed station, and a plan of the desired area is prepared.

4.4 SUMMARY

A contour line is an imaginary line joining parts of equal height above sea level. It is helpful in constructing a physical map and displaying slope contours. Till the last 2–3 decades, all cartography was dependent on the network of contour lines, but at present, new technology has reduced the art of using human hands. As a general rule, there is a vertical gap of 20, 50, and 100 metres between two contour lines, but sometimes it changes according to the slope. Extreme care is required in the construction of the contour line because, besides being very fine, it also represents the actual relief of a region. Before constructing a contour, a surveyor should know the average height of any contour line from sea level. On the basis of that, the contour line is determined. These lines reveal the steep and gentle gradient of the relief. Agriculture is carried out in hilly areas on the basis of the characteristics of the contour lines, and it also helps in determining biodiversity.

Apart from this, it also provides scientific suggestions to prevent soil erosion. In mountainous areas, soil erosion is prevented by doing it according to the contour lines. Thus, contour drawing is an integral part of cartography because displaying relief features through contour lines has been a very useful and popular method. Indian pattern clinometers are simple, small, and light instruments manufactured by the Survey of India that work to measure height and depth in parts of the surface with uneven obstacles. In which angle and natural value are mentioned. The minimum reading angle is 20 minutes, and the tangent value is 0.005. The maximum reading angle can be read up to 22°, and natural tangents are 0.4 only.

Contour lines work to show the shape, height, and slope of a place on the ground. And the contour lines do not intersect each other. Contour lines are mainly expressed in three ways. Intermediate index and supplementary contour lines are also known as plane lines, which are done through an actual survey. In which the value of their height is written. The interval of contour lines indicates the slope of the region. Contour farming, soil erosion, moisture conservation, crop productivity increase, and low-expense farming methods are also being developed. Due to dam formation, a reduction in water velocity and an increase in underground water level occur. Apart from this, the contour line also reveals the relief of the first, second, and third orders, especially the mountain, plateau, plain, waterfall, valley, flat slope, steep slope, great block, lake, river bank, frozen bank, ledge, and other such topographic maps that show the slopes of valleys, hills, and valleys. Which is presented in a horizontal arrangement in the model format of the map? In conclusion, it can be said that contour lines are lines drawn to indicate the heights of land on topographic maps. Whereas Indian pattern clinometers are used especially on the principles of trigonometry in measuring the height of tall trees, pillars, towers, and buildings, or in preliminary survey water development level measurement, etc., hence they are also called tangent clinometers. Thus, a clinometer is a land survey instrument that reveals the height and depth of a geographical feature.

4.5 GLOSSARY

Contour line	A Contour line is an imaginary line Joining Parts of equal height above sea level.
Contour line Interval	Vertical distance between two Contour line.
Relief	Average between minimum and maximum points.
Contour line Interpolation	The art of drawing Contour lines with the help of Spatial heights determined From sea level is called Contour interpolation.
Indian pattern Clinometers	A Survey instrument designed by the Survey of India to measure elevation and Depression Paints.
Eye vane	Instrument hole targeting object in survey area.
Tangent Value	Indian Clinometers minimum measuring value in 0.05 and maximum value is 0.4
Elevation angle	Indian pattern Clinometers elevation & depression Minimum value angle is 20 minutes and Maximum measuring value 22°.

4.6 ANSWER TO THE CHECK YOUR PROGRESS

Q-1. A Contour line is an imaginary line Joining land area of equal height determined by Survey from sea leval.

Q-2. In any Contour map the Interval of Contour Lines in 20,50 and 100 meters vertical distance.

Q-3. Horizontal Distance is also known as Horizontal equivalent H.E.

Q-4. The contour lines are always Natural, they never intersect each other.

Q-5. If the slope of the Surface is height, then the Contour lines are close together and if the slope is Normal, then the contour line are made for apart.

Q-6. Contour lines are meat useful in the relief representation of The slope.

Q-7. Indian Pattern Clinometers to a Survey instrument to determine The Height and Depression of Points from the surface.

Q-8. Indian Pattern clinometers were manufactured by the Survey at India; hence it is named Indian pattern clinometers.

Q-9. with the help of Indian Patten clinometers the height and Depth can be easily determined in the obstructed area.

Q-10. The sliding precisely targets the Points to be observed to be observed in the Survey area.

4.7 REFERENCES

- Sharma, J.P. 2007-2008 Practical Geography, Rastogi Publication Mearth.
- Mishra. P.L. 2010 Practical Geography, Vishv Bharti Publication New Delhi.
- Practical work in geography Part-1 class 11th NCERT, New Delhi.

4.8 TERMINAL QUESTIONS.

(i) Long types Questions

Q-1 . Giving the meaning and definitation of Contour line! Explain the Precautions for the determination at Contour lings?

Q-2. How in the Interpolation of Contour linens and the Determination of height in a contour map? Describe the main features of Contour lines?

Q-3. How is Reliefs Shown in uniform, cone slope, convex slope, Terraced slope, Ridges, Plateau, Conical hill, U shaped valley, water Fall, Riya coast by contour?

Q-4. How are distance Points determined by Indian pattern Clinometer? Prepare a plan for any building in your College Campus?

(ii) Short Types Questions.

Q-1. What are contour lines called?

Q-2. Describe the main Characteristics at contour lines?

Q-3. What do you understand by Contour live Interval?

Q-4. How to Interpolation of Contour lines done?

Q-5. How to the representation of relief by Contour lines, Shown in Concave Slope, Terraced slope?

Q-6. How is the representation at relist by Contour lines shown U. Shaped valley?

Q-7. Describe the Different Parts of the Indian Pattern Clinometers with illustrations?

Q-8. Explain the method of using Indian Pattern Clinometers

Q-9. How in the height at distance Paints determined by The Indian Pattern Clinometers?

Q-10. How is the Indian Pattern Clinometers Named?

(iii) Multiple choice questions.

Q-1. What is shown by the Contour lines?

(i) Points of equal height Air Pressure

(ii) Places of Similar altitude.

(iii) Isotherms

(iv) Isoseismic

Q-2. What is the Standard Interval between Contour lines?

(i) 20, 50, 100 Meters

(ii) 10, 20, 50 Meters

(iii) 50 100, 150 Meters

(iv) 500, 1000 1800 Meters

Q-3. What are the main features of contour lines?

(i) They do not intersect each other.

(ii) Contour lines Join Parts of equal Height above sea level.

(iii) Representation at Relief is shown by contour line.

(iv) All at the above.

Q-4. Contour lines are located in Steep Slope?

(i) Close-Close.

(ii) Far away.

(iii) Near and far away.

(iv) None of the above.

Q-5. Contour lines are located on gentle slope?

(i) Near and Far away.

(ii) far away

(iii) Close-Close

(iv) All at the above.

Q-6. The shape at the Contour lines in formed to show the like?

(i) Spherical

(ii) Circular Shape

(iii) Triangular

(iv) Rectangular-

Q**-7. What is the anther name of Indian Pattern Clinometers?**

(i) Tangent Clinometers

(ii) Dumpy Level

(iii) Telescope Elided

(iv) All of the above

Q-8. The Maximum and minimum tangent values in the Indian Pattern clinometers are...?

(i) 0.005 to 0.4

(ii) 0.005 to 0.5

(iii) 0.005 to 0.3

(iv) 0.003 to 0.6

Q-9. The Minimum & maximum angle on the India Pattern clinometers are....?

(i) 20 minutes to 10^0

(ii) 20 Minutes to 22°

(iii) 10 Minutes to 30^0

(iv) 20 Minutes to 40^0

Q**-10. Indian Pattern clinometers in used.**

(i) Angle of elevation

(ii) Angle of Depression

(iii) Angle of Elevation & Depression

(iv) all of the above

Answer's- 1- ii 2- i 3- iv 4. i 5- ii 6- ii 7- i 8-i 9- ii 10- iii

UTTARAKHAND OPEN UNIVERSITY

Teenpani Bypass Road, Behind Transport Nagar,
Haldwani- 263139, Nainital (Uttarakhand)
Phone: 05946-261122, 261123; Fax No. 05946-264232
Website: www.uou.ac.in; e-mail: info@uou.ac.in
Toll Free No.: 1800 180 4025