MIT | School of Distance Education

Chapter 8

Data Processing and Analysis

Chapter Index

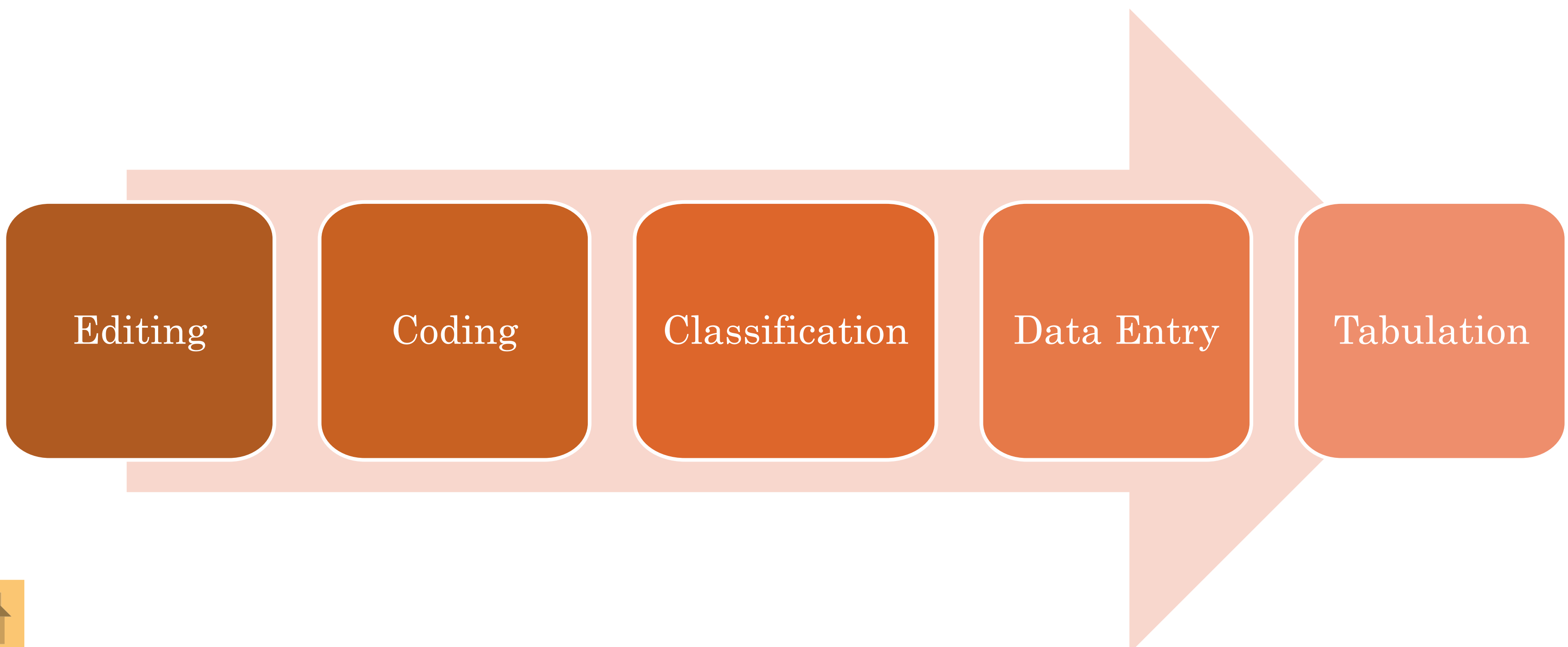
Learning Objectives

- Explain the concept of data processing
- Describe the concept of data analysis
- Discuss the measures of central tendency
- Explain the measures of skewness
- Discuss the measures of relationship
- Describe the various charts used in data analysis

The Concept of Data Processing

The processing is a process of converting raw data (quantitative or qualitative) into a form, which is fit for analysis.

Various steps involved in data processing



Editing

Editing refers to reviewing the collected data to check whether it is valid or not.

Data is examined to detect errors and omission. Errors are corrected, omitted data is filled in, and data is prepared for further processing.

The editor is responsible for ensuring that the data is accurate, uniform, as complete as possible and acceptable for tabulation.

Editing helps in filtering ambiguous information that can create a problem at the time of data analysis.

Coding

Coding is the process of providing some codes to the data in the form of symbols, characters and numbers which helps the researcher in interpreting the data and deriving accurate results.

Precoded data: The data that is already coded.

Postcoded data: The data that is coded at the time of data processing.

Types of questions that a questionnaire may contain

Interval-scale questions
It is any range of values that have a relevant mathematical difference but no true zero.
A temperature value is an interval scale question because degrees are interval measurements.

Open-ended questions
These questions are those which require more thought and more than a simple one-word answer. The data collected through open-ended questions is an example of postcoded data.

Closed-ended questions
These questions are those for which a researcher provides respondents with options from which to choose a response.

Precoded data has certain advantages over postcoded data:

- It is easier to code.
- It reduces the effort in data processing.
- It leads to fewer chances of human error during data processing.

Classification

Classification refers to categorising the coded questions into different segments as per their relevance.

Questionnaire can be classified into:

Qualitative questions
The classification of qualitative questions is called statistics of attributes. These attributes cannot be measured directly in numbers. However, qualitative attributes can be quantified.
Example: Honesty and attitude of the respondents

Quantitative questions
The classification of quantitative questions is called statistics of variables. These variables can be expressed in numeric form, such as demographic factors including age and income.
Example: In the class interval 25-35, 25 is the lower limit and 35 is the upper limit.

Class intervals can be

Inclusive class intervals

If the value of the upper limit is included in the class magnitude, it is an inclusive class interval.

Example: The value 35 would be included in the inclusive class 25-35

Exclusive class intervals

If the value of the upper limit is not included in the class magnitude, then it is an exclusive class interval.

Example: The value 35 would not be included in the class 25-35 but it would be included in group 35-45.

Frequency is the number of occurrences of a repeating event per unit of time.

Example

- The number of respondents in each age group is shown in the below table:

Age Group (Class Interval)	Number of Respondents
25-35	10
35-45	4
45-55	7
55-65	2

- In the above table, 10 is the frequency of the class interval 25-35.
- When class intervals and frequencies are represented in a tabular form, as in Table 1, such a representation is known as frequency distribution.

Data Entry

After classifying data, the researcher enters data in the computer. If wrong data is entered, then the result would be inaccurate.

There are various statistical or database management software for data entry, such as:

- Bio Medical Data Package (BMDP)
- Statistical Programming Language (S-PLUS)
- Statistical Analysis System (SAS)
- Statistical Package for Social Sciences (SPSS)

Tabulation

Tabulation refers to presenting data in the form of a table so that it can be easily analysed.

There are three types of frequencies

Absolute frequency
It is the exact frequency given by the respondents.

Relative frequency
It s calculated with relation to the frequency of the other class intervals. It is the percentage of all respondents who have given a particular response.

Cumulative frequency
It is the percentage of all respondents who have given a response equal or less than a particular value.

Here, two variables can be analysed at a time.

Two-way frequency distribution (Cross tabulation)

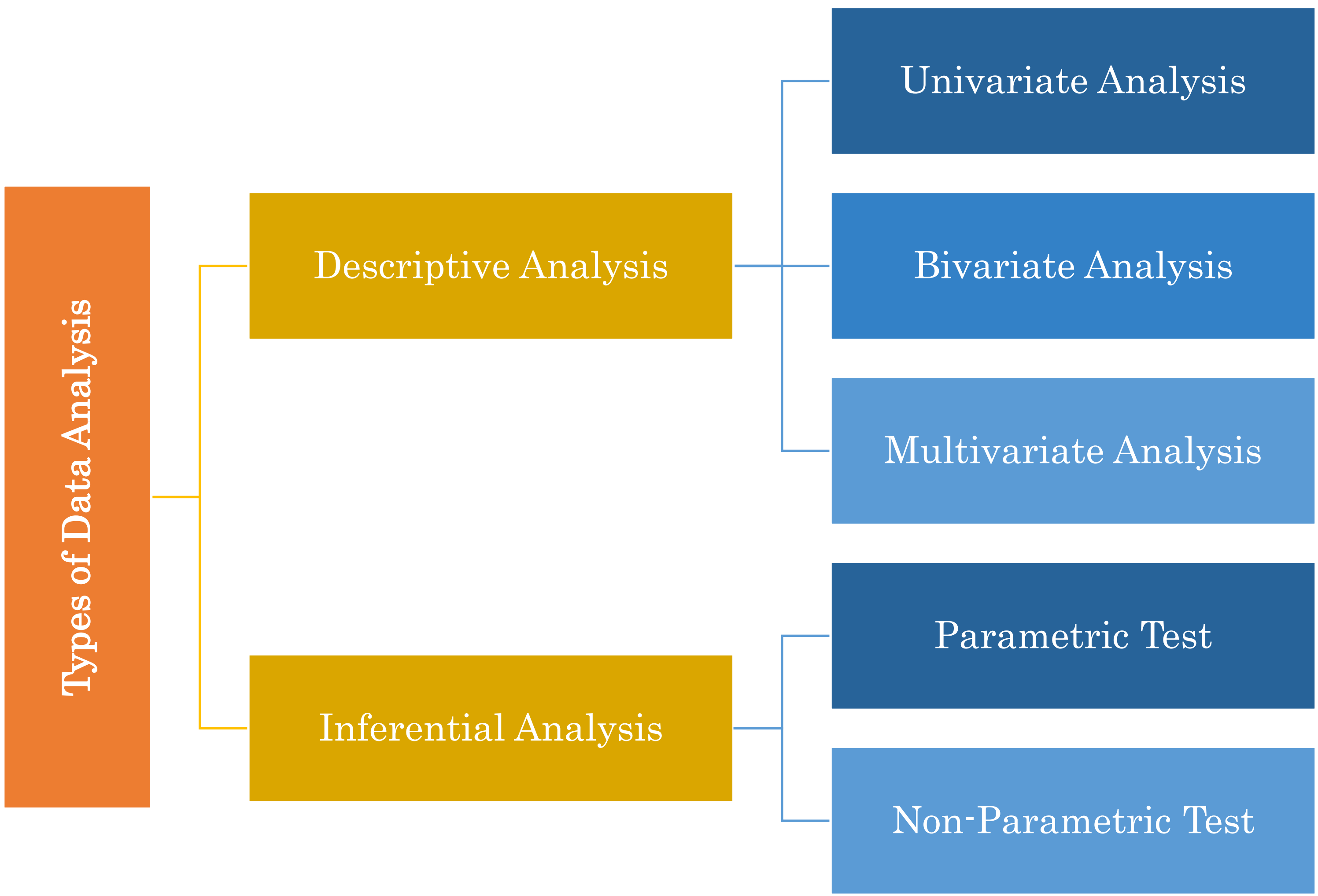
The two types of frequency distributions, which can be put into a tabular form are:

One-way frequency distribution

In this case, a single variable is analysed.

Example of the one-way frequency distribution			
Age Group (Class Interval)	**Number of Persons (Frequency or Absolute Frequency)**	**Relative Frequency**	**Cumulative Frequency**
20-30	10	17.86	17.86
30-40	14	25.00	42.86
40-50	20	35.71	78.57
50 and above	12	21.43	100.00
Total	56	100	100

The Concept of Data Analysis



Descriptive Analysis

In this type of data analysis, the distribution patterns and characteristics of different types of variables are analysed.

Univariate Analysis

This analysis studies a single variable.

Examples include measures of central tendency, dispersion and skewness.

Bivariate Analysis

Here, two variables are studied. One variable can be classified as independent and the other as dependent.

Examples are rank correlation, simple correlation and simple regression.

Multivariate Analysis

Here, more than two variables are studied among which, there can be more than two independent variables and more than one dependent variable.

Examples include multiple correlations and regressions.

Inferential analysis

In this type of data analysis, significance tests are used to check the validity of a hypothesis for studying a problem.

Parametric tests

These tests make assumptions about the parameters of the population from which a sample is derived.

Examples of parametric tests include z-test and t-test.

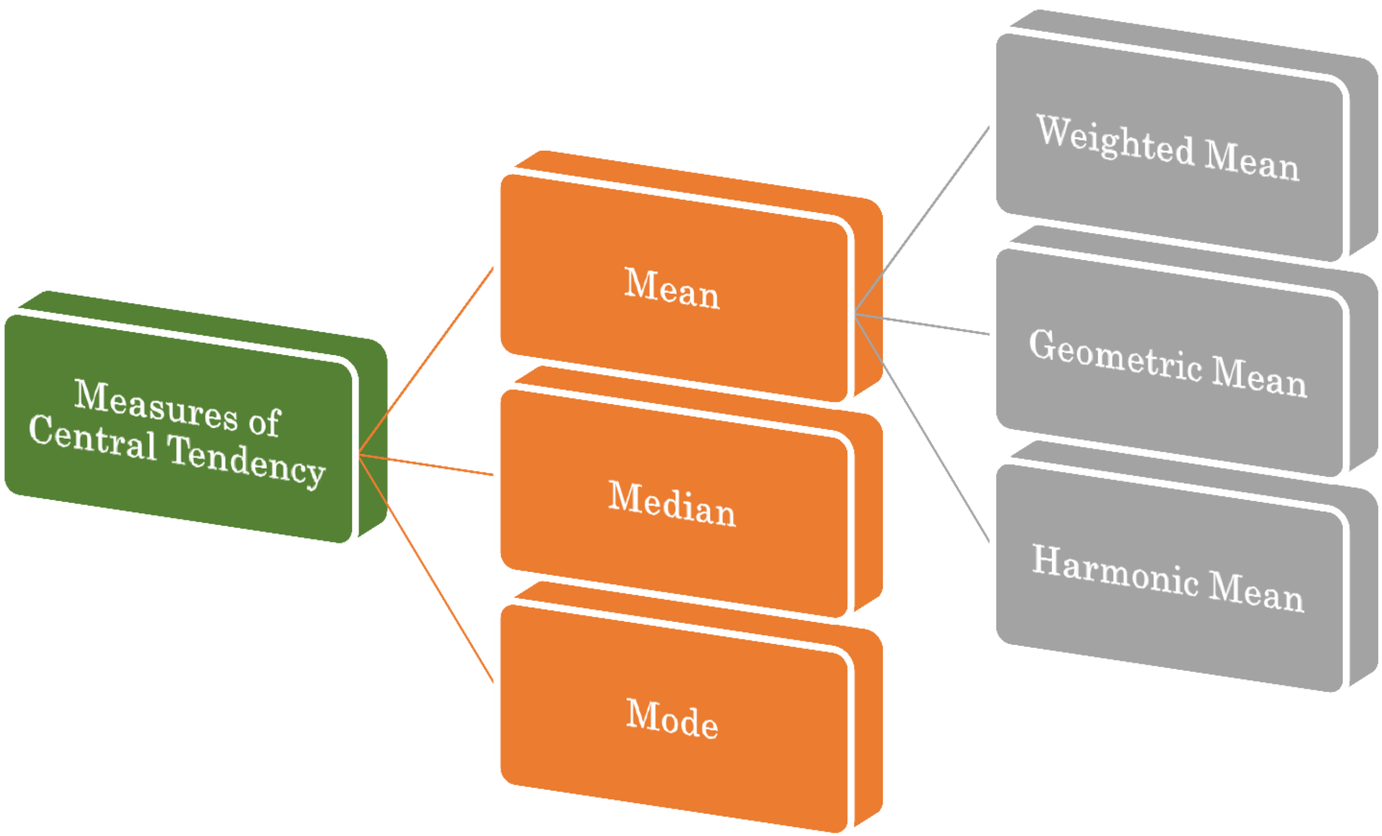
Non-parametric tests

These tests do not make any assumptions about the parameters of the population from which the sample is derived.

An example is the Kruskal Wallis test.

Measures of Central Tendency

The measure of central tendency give a central value that represents the large chunk of data analysed.



Mean

- Mean (arithmetic mean) represents the value calculated after dividing the sum of observations by the total number of observations (n) taken.
- Formula: Mean (X) = $\bar{X} = \Sigma X_i/n$

  where ΣX_i sum of all the observations & n is Total number of observations

Example: The weight of five friends is given in the below table. Find the average weight of a group of five friends.

People	Weight (kg)
Jenny	35
Robert	40
Ella	34
Andy	39
Eliza	42

The average weight of five friends is:

$\bar{X} = \Sigma X_i/n$

$= (35 + 40 + 34 + 39 + 42)/5$

$= 190/5$

$= 38$ kg

Therefore, the average weight of five friends is 38 kg.

Weighted mean

This mean is calculated after considering the weight attached to each item.

Formula: Weighted Mean ($\overline{X_w}$) = Σ WiXi/wi

where X_i = Value of the ith item;

W_i = Weight assigned to the ith item;

w_i = Number of weights assigned

Example: A school grades its students by using weighted mean scores as follows: 15% weightage is assigned for homework, 15% weightage is assigned for extracurricular activities, and 70% weightage is assigned for the examination. Aditya scored 60 marks, 70 marks and 55 marks for homework, extracurricular activities and in examination respectively. Find the weighted score of Aditya if the total score is 100.

Weighted Mean ($\overline{X_w}$) = (0.15 × 60) + (0.15 × 70) + (0.70 × 55)

= 9 +10.5+ 38.5

= 58

Geometric mean

Geometric mean represents the nth root of the product of all the values or observations involved in a research.

Formula: $\overline{X_g} = \sqrt[n]{(X_1)(X_2)(X_3)\ldots(X_n)}$

where X_1, X_2, …………, X_n are the n observations in the data set

n = Number of observations

Example: You want to calculate the geometric mean of four observations: 10, 12, 10 and 11.

$$\overline{X_g} = \sqrt[4]{(X_1)(X_2)(X_3)(X_4)}$$
$$= \sqrt[4]{(10)(12)(10)(11)}$$
$$= 10.718$$

Therefore, the geometric mean of four observations is 10.7 years.

Harmonic mean

Harmonic mean refers to reciprocal of the average of the reciprocals of the values in a data series (or observations).

Formula: Harmonic mean Rec. $[(\text{Rec.}X_1 + \text{Rec.}X_2 + \cdots \ldots \ldots \ldots . + \text{Rec.}Xn)/n]$

where $\text{Rec.}X_1, \text{Rec.}X_2, \ldots \text{Rec.}Xn$ are Reciprocal of Observations 1, 2, ...n, respectively.

n = Number of observations

Example: Calculate the harmonic mean of four observations: 10, 12, 10 and 11.

$\overline{X_H} = \text{Rec.}[(\text{Rec.}X_1 + \text{Rec.}X_2 + \text{Rec.}X_3 + \text{Rec.}Xn)/n]$

$= \text{Rec.}\ [(\frac{1}{10} + \frac{1}{12} + \frac{1}{10} + \frac{1}{11})/4]$

$= \text{Rec.}[(\frac{247}{660})/4)$

=10.68

Therefore, the harmonic mean of the four observations is 10.7 years.

Median

Median is defined as a central or mid value of a dataset.

Formula: Let n = Number of observation. First arrange the dataset in the ascending or descending order, then calculate median using the formula:

- If n is an odd number then

  Median = $\left(\frac{n+1}{2}\right)^{th}$ observation

- If n is an even number then

  Median = Value of { [$(n/2)^{th}$ observation + $((n+1)/2)^{th}$ observation] / 2 }

Example: A group of 17 people gave the following ratings to a book on a 5-pointer scale (where 1 is the lowest rating and 5 is the highest rating): 2, 5, 3, 4, 1, 5, 4, 3, 1, 2, 5, 4, 3, 2, 1, 5, 4. Calculate median.

- First arrange the values in ascending order:

  1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5

- Since number of observations is an odd number the formula will be used

  Median = $\left(\frac{n+1}{2}\right)^{th}$ observation = $\left(\frac{17+1}{2}\right)^{th}$ observation = 9^{th} Observation

  Median = 3

Example 2: If a group of 20 people gave their ratings to a movie on a 5-point scale as:

2, 5, 3, 4, 1, 5, 4, 3, 1, 2, 5, 4, 3, 2, 1, 5, 4, 1, 2, 3

Where, 1 is the lowest rating and 5 is the highest rating

- First arrange the values in ascending order:
  1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5
- Here, median is the average of middle two values, i.e., values at 10^{th} and 11^{th} positions. This is calculated as:
  Median = (3 + 3)/2 = 3

Mode

Mode refers to the value that has the highest frequency in a data series.

According to **Croxton and Cowden**, the mode of a distribution is value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values.

Example: Suppose the marks of five friends in a science paper are 70, 90, 50, 70, and 30. Find the mode of their marks.

- You need to find the highest frequency of the present data to calculate mode.
- Here, the number having the highest frequency is 70 as it occurs two times; therefore, the mode of students' marks is 70.

Mode is not considered a true measure of central tendency because of three reasons:

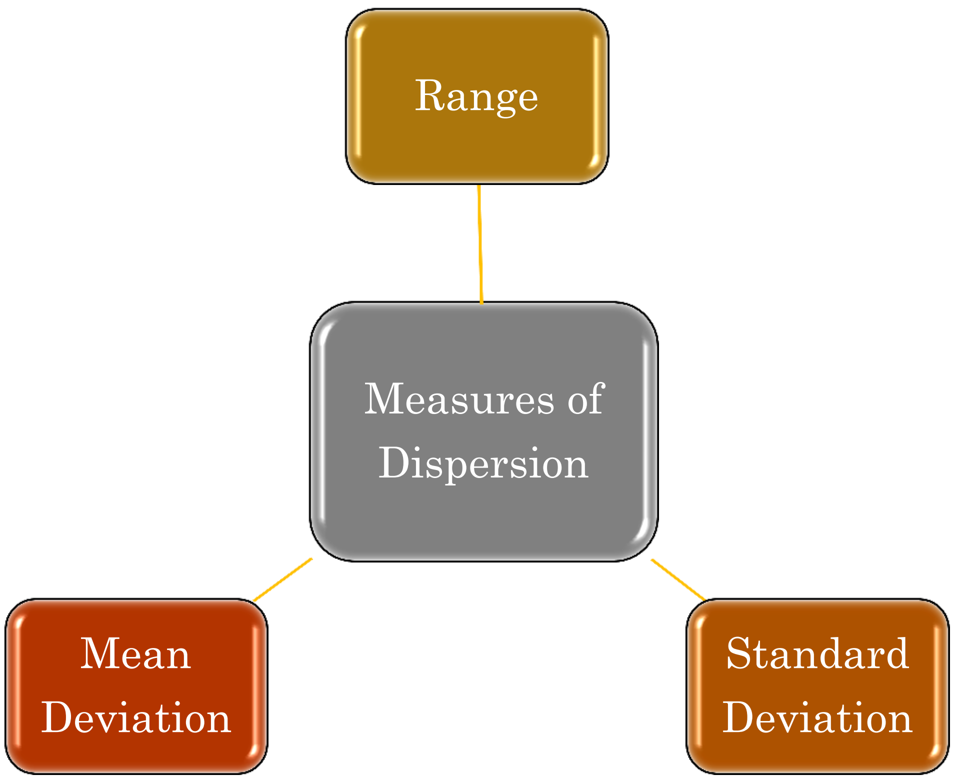
- It is not necessary that one data series has only one mode because many numbers in the data series can have the highest frequency.
- Mode does not consider all the frequencies to arrive at the central value of the data series. Therefore, the results of mode are not reliable.
- It is possible that a series has observations that occur only once. In such cases, mode does not exist.

Types of Variables and Measures of Central Tendency

Types of Variables	Best Measure of Central Tendency
Nominal	Mode
Ordinal	Median and Mode
Interval/Ratio (not skewed)	Mean, Median and Mode
Interval/Ratio (skewed)	Median and Mode
* For skewed data, median is better than mean	

Measures of Dispersion

The measures of dispersion can be used to study the dispersed values near the mean value.



Range

Range represents the difference between the highest value and the lowest value in a data series.

Formula: Range = (Highest value of data series – Lowest value of data series)

Example: A group of 17 people rated a book on a 5-pointer scale, where 1 is the lowest rating and 5 is the highest rating. The rating given by the 17 people is as follows:

2, 5, 3, 4, 1, 5, 4, 3, 1, 2, 5, 4, 3, 2, 1, 5, 4

Highest value of data series = 5

Lowest value of data series = 1

Therefore, range= (Highest value of data series – lowest value of data series)

Range = (5 – 1) = 4

Mean Deviation

Mean deviation represents the extent of deviation of values from the mean. It is used to measure variability across a data series.

Formula: Mean Deviation (MD) = $\Sigma|X_i - \bar{X}|/n$

Where X_i = Individual observation; $\bar{X}$ = Mean/Median/Mode;

n = Number of observations

The coefficient of MD refers to the relative measure of dispersion that can be calculated by dividing MD with mean/median/mode.

Formula: Coefficient of MD = $MD/\bar{X}$

Where MD = Mean deviation; $\bar{X}$ = Mean/Median/Mode

Example: The average weight of five friends is given in the below table.

People	Weight (kg)	$\lvert X_i - \bar{X} \rvert$
Jenny	35	$\lvert 35 - 38 \rvert = 3$
Robert	40	$\lvert 40 - 38 \rvert = 2$
Ella	34	$\lvert 34 - 38 \rvert = 4$
Andy	39	$\lvert 39 - 38 \rvert = 1$
Eliza	42	$\lvert 42 - 38 \rvert = 4$
Total		14

$$\bar{X} = \frac{35 + 40 + 34 + 39 + 42}{5} = 38$$

Mean Deviation (MD) = $\Sigma \lvert X_i - \bar{X} \rvert / n = 14/5 = 2.8$

Coefficient of MD = $MD/\bar{X} = 2.8/38 = 0.074$

Therefore, the weight of all friends is dispersed more or less by 2.8 kg from the average weight.

The relative measure of weight is 0.074.

Standard Deviation

Standard Deviation (SD) is used to calculate the scattering of values in a given dataset. The symbol used to represent standard deviation is sigma (σ). (SD) is the square root of variance of a data series.

Formula:

- SD of population, $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n}}$
- SD of sample, $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n-1}}$
- Population variance = σ^2 and Sample variance = S^2
- If the observations are grouped into a frequency table, than the formula of SD and variance change as follow:

  $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{x})^2 f_i}{n}}$ where $\bar{X} = \frac{\sum_{i=1}^{n} X_i f_i}{\sum_{i=1}^{n} f_i}$; $n=\sum_{i=1}^{n} f_i$
- The coefficient of SD = $\frac{\sigma}{\bar{X}}$

Example: The average weight of five friends is given in the below table. Calculate the standard deviation

People	Weight (kg) (X_i)	($X_i - \bar{X}$)	$(X_i - \bar{X})^2$
Jenny	35	− 3	9
Robert	40	2	4
Ella	34	− 4	16
Andy	39	1	1
Eliza	42	4	16
Total			$\sum(X_i - \bar{X})^2 = 46$

$$\bar{X} = \frac{35 + 40 + 34 + 39 + 42}{5} = 38$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n}} = \sqrt{\frac{46}{5}} = 3.033$$

Coefficient of SD = SD/$\bar{X}$ = 3.03/38 = 0.098

Measures of Skewness

The measure of skewness is used to study the shape of a curve that can be drawn by plotting the data of a frequency distribution.

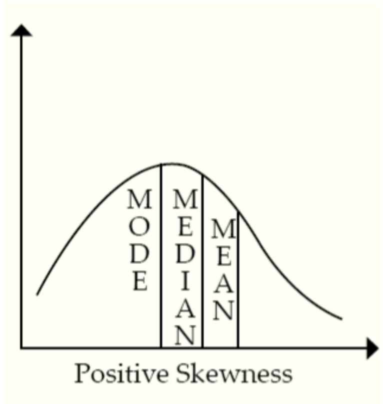
The measure of skewness is used when the concentration of values of a data series is more on a single side that is either positive or negative.

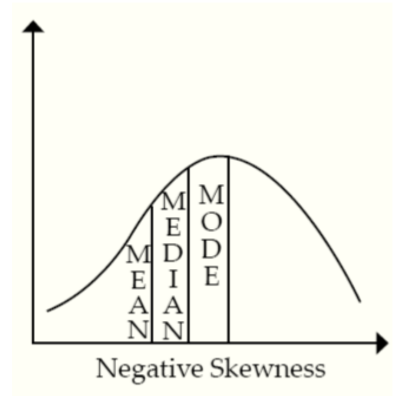
Formula

Skewness = $\bar{X} - Z$; Coefficient of Skewness, $S_k = \frac{\bar{X} - Z}{\sigma}$

For a moderately skewed, if there is more than one mode or if there is no mode, then skewness is calculated using the method of Moments, that is, Skewness = $3(\bar{X} - M)$

Skewness can be classified as positive skewness and negative skewness.





Positive skewness implies that the concentration of values is on the right side of the curve

Negative skewness implies that the concentration of values is on the left side of the curve.

In positive skewness, the order of the three measures are as follows Mean > Median > Mode

In negative skewness, the order of the three measures are as follows Mean < Median < Mode

Example: Suppose you want to calculate the skewness and the coefficient of skewness of the data given in Table:

People	Age (Years)	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
Jenny	18	0.2	0.04
Robert	17	-0.8	0.64
Ella	18	0.2	0.04
Andy	17	-0.8	0.64
Eliza	19	1.2	1.44
Total	$\sum X_i = 89$		$\sum(X_i - \bar{X})^2 = 2.80$

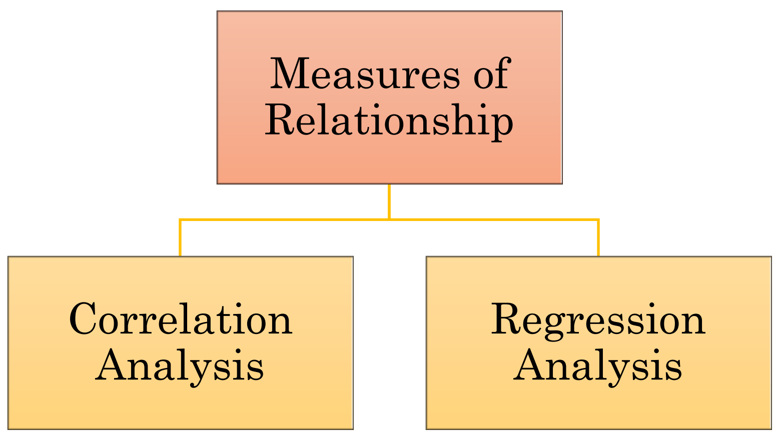
- Mean of age, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = 89/5 = 17.8$
- Median, M = Value of ((n+1)/2)th value = ((5+1)/2)th value = 3rd observation = 18
- Since there are two modes (17 & 18), it is not considered.
- Hence, Skewness = $3(\bar{X} - M) = 3(17.8 - 18) = 0.6$
- SD of age, $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n}} = \sqrt{\frac{2.80}{5}} = 0.75$
- The coefficient of skewness = 0.6/0.75 = 0.8

Measures of Relationship

The measures of relationship study the relationship between two or more variables in a given data series.

When the relationship between two variables in a population is studied, it is known as bivariate population.

When the relationship between more than two variables in a population is studied, it is known as multivariate population.



Correlation Analysis

Correlation analysis is used to study the association between different types of variables.

Tools used to study the correlation pattern between variables are:

- Rank Correlation
- Simple Correlation

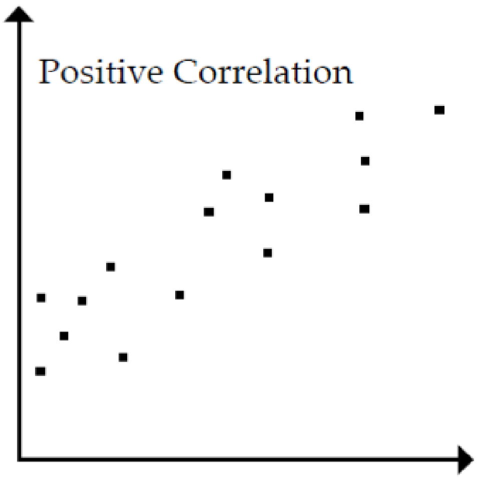
Rank Correlation

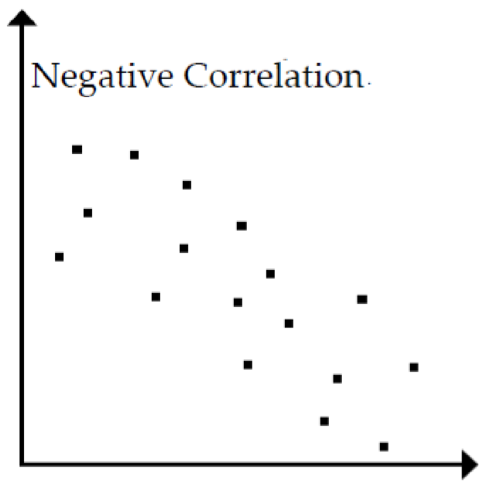
- It refers to the correlation between two data series in which the data is ranked.
- Generally, it is found when the data is qualitative in nature.
- It calculates the degree of relationship between two types of variables.
- It is calculated as $\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$
- Where, d_i = Difference between the individual / i^{th} pair of variables; n = Number of pairs of observations

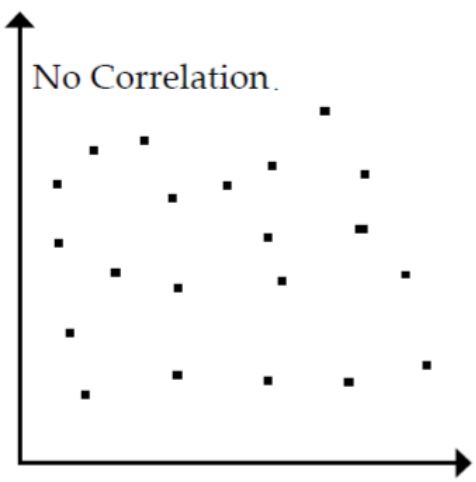
Simple correlation

It is used to find the degree of linear relationship between two variables. It is also known as Karl Pearson's coefficient of correlation.

Simple Correlation can be of three types







The value of the correlation coefficient lies between a range of –1 and +1.

If the value of the correlation coefficient is close to –1 and the sample size is sufficiently large, then there is **a strong negative correlation** between two variables.

If the value of the correlation coefficient is close to +1 and the sample size is sufficiently large, then there is a **strong positive correlation** between two variables.

If the correlation coefficient is not close to –1 or + 1 and the sample size is sufficiently large, then there is **weak correlation between** two variables.

Different formulae to calculate correlation

1. $r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_x S_y}$

2. $r = \frac{Cov(X,Y)}{SD_x SD_y}$

3. $r = \frac{(n\sum_{i=1}^{n} X_i Y_i) - (\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$

Where, X_i = i[th] value of X variable; $\bar{X}$ = Mean of X variable; Y_i = i[th] value of Y variable; $\bar{Y}$ = Mean of Y variable; n = Number of pairs of observations; S_x = Standard deviation of X; S_y = Standard deviation of Y

Example: To study the correlation between the age and weight of a group of people to find out the relation between the two.

Number of observations	Age (X_i)	Weight(Y_i)	X_i^2	Y_i^2	X_iY_i
1	18	35	324	1225	630
2	20	38	400	1444	760
3	25	50	625	2500	1250
4	30	65	900	4225	1950
5	35	70	1225	4900	2450
6	24	50	576	2500	1200
7	17	35	289	1225	595
8	16	39	256	1521	624
9	49	76	2401	5776	3724
10	45	72	2025	5184	3240
11	50	85	2500	7225	4250
12	18	32	324	1024	576
13	20	34	400	1156	680
14	25	57	625	3249	1425
15	24	50	576	2500	1200
16	17	35	289	1225	595
17	16	39	256	1521	624
18	23	44	529	1936	1012
19	22	45	484	2025	990
20	34	60	1156	3600	2040
21	36	65	1296	4225	2340
22	31	63	961	3969	1953
23	43	70	1849	4900	3010
24	44	72	1936	5184	3168
25	16	35	256	1225	560
Total	698	1316	22458	75464	40846

$$r = \frac{(n\sum_{i=1}^{n} X_iY_i)-(\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

$$r = \frac{(25\times40846)-(698\times1316)}{\sqrt{[25\times22458 - 698\times698][25\times75464 - 1316\times1316]}}$$

$$r = 0.96$$

Regression Analysis

Regression is one step ahead of correlation in identification of relationship between two variables because regression allows for prediction of values within the given data range.

- The regression equation can be written as Y = α + βX, where Y represents scores on Y variable, X represents scores on X variable, α represents regression constant in the sample, β represents regression coefficient in the sample.
- The variable Y is generally termed as dependent or criterion variable and the variable X is termed as independent or predictor variable.
- Regression equation is used to generally predict the values of Y based on the values of X.

α and β are calculated as

$$\beta = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \qquad \alpha = \frac{1}{n}\left[\sum_{i=1}^{n} Y_i - \beta \sum_{i=1}^{n} X_i\right]$$

Example: The data of number of customers and monthly sales for 10 number of observations (N) as shown in the below Table:

S. No.	No. of Consumers (X) (in '00)	Monthly Sales (Y) (in '000)	X_iY_i	X_i^2
1	2	12	24	4
2	3.4	6	20.4	11.6
3	6.2	7	43.4	38.4
4	7.6	11	83.6	57.8
5	6.5	13	84.5	42.3
6	8.2	33	270.6	67.2
7	7.6	31	235.6	57.8
8	9.3	22	204.6	86.5
9	3.1	36	111.6	9.6
10	8.1	24	194.4	65.6
Total	**62**	**195**	**1272.7**	**440.8**

The regression equation can be written as Y = α + βX

$$\beta = \frac{n\sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

$$= \frac{10\times1272.7 - (62)(195)}{10\times440.7 - (62)(62)} = 1.1314$$

$$\alpha = \frac{1}{n}\left[\sum_{i=1}^{n} Y_i \;-\; \beta\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{10}\,[195 - 1.1314 \times 62] = 12.485$$

Thus, the regression equation for the above data is given as: Y = 12.485 + 1.1314X

With the equation Y = 12.485 + 1.1314X, the values of Y (monthly sales) can be computed for any given value of X (no. of customers) as depicted in the below table.

S. No.	No. of Consumers (X) (in '00)	Y=12.485+1.1314X	Monthly Sales (Y) (in '000)
1	2.0	14.75	(12.485 + 1.1314×2.0)
2	3.4	16.33	(12.485 + 1.1314×3.4)
3	6.2	19.50	(12.485 + 1.1314×6.2)
4	7.6	21.08	(12.485 + 1.1314×7.6)
5	6.5	19.84	(12.485 + 1.1314×6.5)
6	8.2	21.76	(12.485 + 1.1314×8.2)
7	7.6	21.08	(12.485 + 1.1314×7.6)
8	9.3	23.01	(12.485 + 1.1314×9.3)
9	3.1	15.99	(12.485 + 1.1314×3.1)
10	8.1	21.65	(12.485 + 1.1314×8.1)
Total	**62.0**	**195.00**	

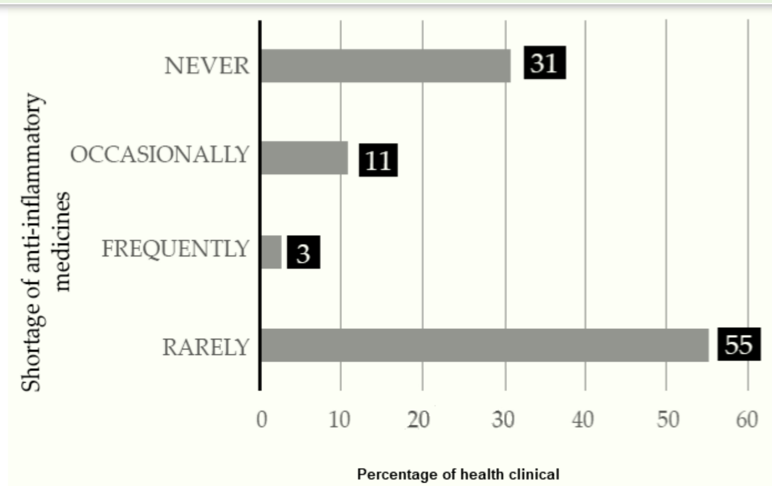
Different Charts Used in Data Analysis

The most frequently used graphs and charts include the following:

Bar Chart

A bar chart represents categorical data with the help of rectangular bars, plotted vertically or horizontally. The heights or lengths of rectangular bars are proportional to the values represented by them.

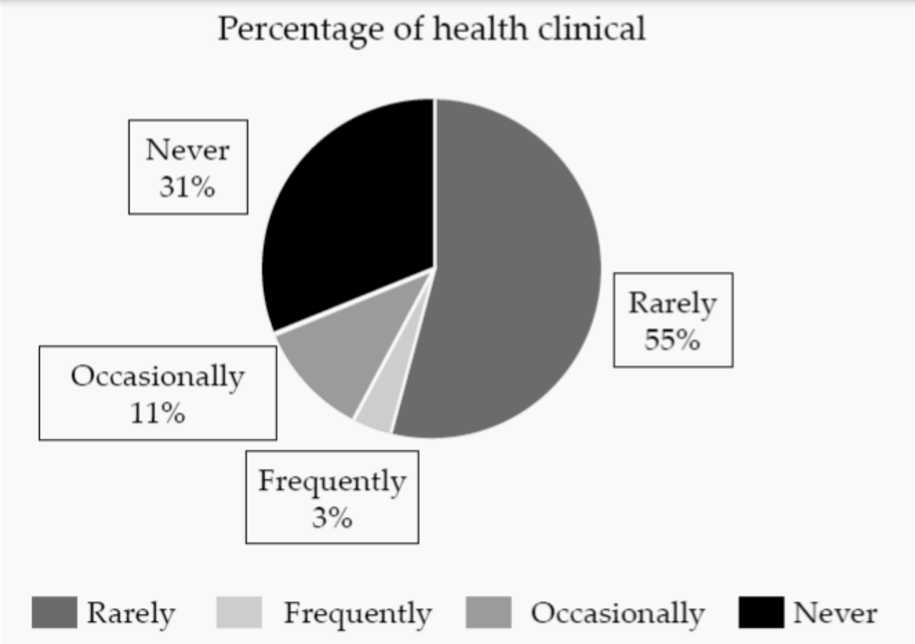
Relative Frequency of Shortages of Anti-inflammatory Medicines in Rural Health Organisations in Bar Chart



Pie Chart

- A pie chart is a circular statistical graphic, segregated into different segments to illustrate the numerical proportions/relative frequency of a number of items.
- The arc length of each segment shows the proportionate quantity represented by it.

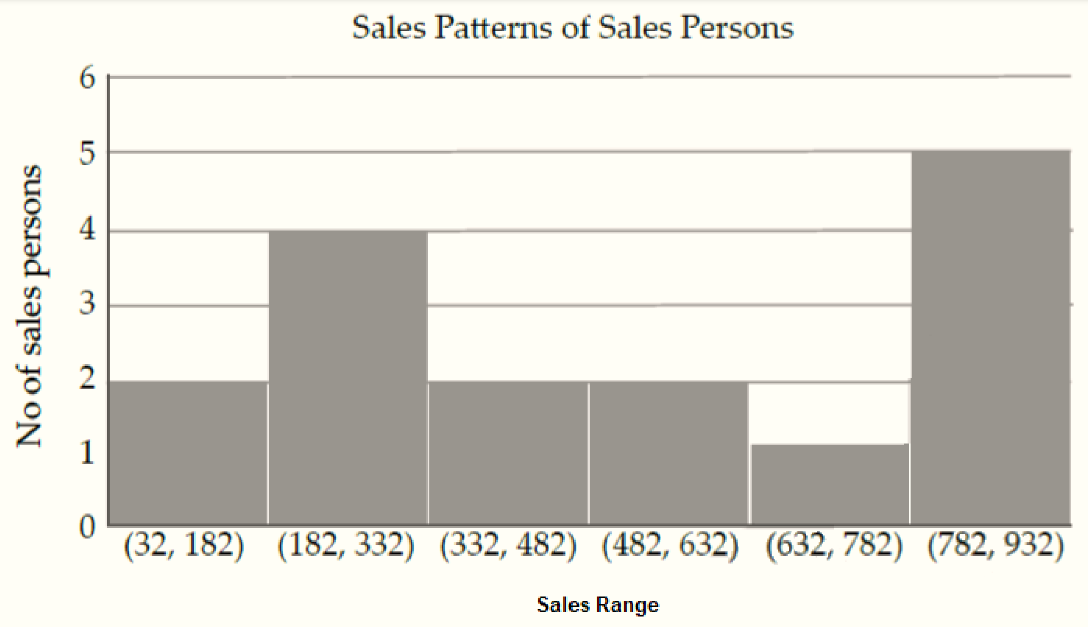
Relative Frequency of Shortages of Anti-inflammatory Medicines in Rural Health Organisations in Pie Chart



Histogram

A histogram is an accurate representation of the probability distribution of a continuous data variable grouped into bins.

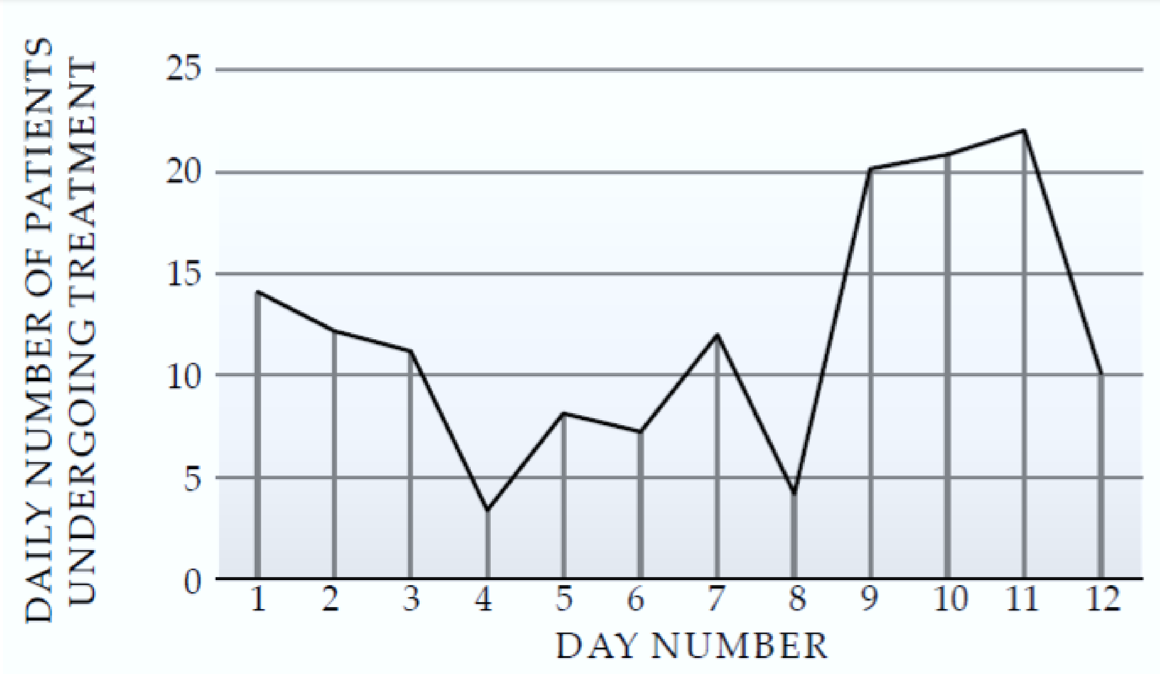
Absolute Frequency of Sales Effected by Different Sales Persons in a Month (n=60)



Line Graph

A line graph or a line chart is generally used to visualize the value of a particular variable over time. They are useful to show the trend of numerical data over a period of time.

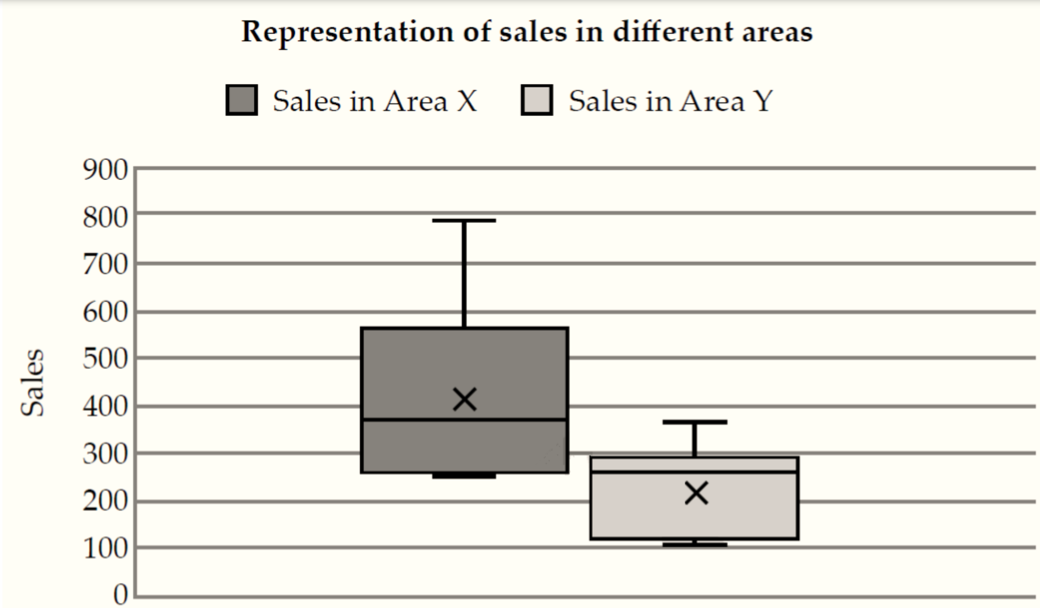
Daily Number of Patients Being Treated at the Rural Health Organisations in District Y in Line Chart



Box and Whisker Plot

This is a method of graphically representing different groups of numerical data through their quartiles.

Sales Patterns of Food Grains Effected in Area X and Area Y



Thank You