Block 4
Composite Index Numbers and Qualitative Data

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 10 CONSTRUCTION OF COMPOSITE INDEX IN SOCIAL SCIENCES

Structure

10.0 Objectives

10.1 Introduction

10.2 Composite Index: The Concept

10.3 Steps in Constructing Composite Index

10.4 Dealing with Missing Values and Outliers

10.5 Methods to Construct Composite Index

- 10.5.1 Simple Ranking Method
- 10.5.2 Indices Method
- 10.5.3 Mean Standardisation Method
- 10.5.4 Range Equalisation Method

10.6 Illustrations of Composite Indices

- 10.6.1 Physical Quality of Life Index (PQLI)
- 10.6.2 Human Development Index (HDI)
- 10.6.3 Gender Development Index (GDI)

10.7 Merits and Limitations of Composite Index

10.8 Let Us Sum Up

10.9 Exercises

10.10 References

10.11 Answers or Hints to Check Your Progress Exercises

10.0 OBJECTIVES

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

- state the concept of composite index;
- describe the process of constructing the composite index;
- explain the various methods of compiling a composite index;
- discuss the merits and limitations of composite index; and
- learn how to interpret the results derived from composite indices.

10.1 INTRODUCTION

In social sciences research, many a times the complex social and economic issues like child deprivation, food security, human well-being, human development etc. are difficult to measure in terms of single variable. The reason being that such issues have several dimensions and indicators. For

example, it is difficult to explain the status of development of a district in terms of a single variable because development is reflected in terms of several indicators. Some of such variables/indicators are quantitative type while others are of qualitative nature. In such situations Composite index plays an important role to express the single value of several interdependent or dependent variables. Further, the composite index makes it possible to compare the performance among region/states or districts etc. Hence, composite indexes are being recognised as useful tool for policy analysis. Composite indexes can also be resorted to make comparison among different regions/sectors where wide range of variables is used.

In this Unit, we shall discuss the concept of composite indices, the process of their construction, various methods, their uses, and limitations. The study of the unit will also enable you to learn how to interpret the results. Let us begin with the concept of composite index.

10.2 COMPOSITE INDEX: THE CONCEPT

A composite index is an expression in a single score made by combination of different scores to measure a given variable or a group of variables. It expresses quantity or place (position) of multi facet aspects of a concept. The UNDP (2005) has explained that 'a composite index expresses a quantity or a position on a scale of qualitative multi-faceted aspects….. which is relevant for information of the society'. An index can be a combination of independent indicators, or the average of a set of selected indicators. The index may represent specific concept or highlight specific sector or areas like status of dalit (dalit deprivation index), food situation (food security/ insecurity index), status of child (child deprivation/development index), quality of human development (human development index), etc.

The index that we construct is the outcome of some unidirectional variables or indicators. If an index is constructed by taking positive indicators, the higher value of index implies higher development and lower values imply lower development. For example, in case of index related to child development, if the variables are positive directional, the final index can be termed as 'child development index'. On the other hand, if the variables are negative directional, it is called 'Child Deprivation Index'. Let us take an example that child mortality is a negatively directional variable whereas percentage of children immunised is a positively directional variable. Such index is very useful in case of qualitative data. An index is more robust than a single indicator or variable.

The choice of indicator is a big challenge for researchers. The major issues in identifying measurable indicators are: whether data are available or not, whether we have reliable data, whether the data are cross section or time series, and minimisation of double counting arising from overlap or redundancy. Again if the variables to be chosen are easy to understand, they are more acceptable to a wider audience.

10.3 STEPS IN CONSTRUCTING COMPOSITE INDEX

The following steps are required to be followed in construction of composite index:

Steps	Why it is Needed?
1. Theoretical framework A theoretical framework need to be developed because it provides the basis for the selection and combination of variables into a meaningful composite indicator under a fitness-for-purpose principle (involvement of experts and stakeholders is envisaged at this step).	To get clear understanding and definition of the multidimensional phenomenon to be measured. To structure the various sub-groups of the phenomenon (if needed). To compile a list of selection criteria for the underlying variables, e.g., input, output, process.
2. Data selection Indicators should be selected on the analytical soundness, measurability, country coverage, and relevance of the indicators to the phenomenon being measured and relationship to each other. The use of proxy variables should be considered when data are scarce (involvement of experts and stakeholders is envisaged at this stage).	To check the quality of the available indicators. To discuss the strengths and weaknesses of each selected indicator. To create a summary table on data characteristics, e.g., availability (across country, time), source, type (hard, soft or input, output, process).
3. Imputation of missing data Consideration should be given to different approaches for imputing missing values. Extreme values should be examined as they can become unintended benchmarks.	To estimate missing values. To provide a measure of the reliability of each imputed value, so as to assess the impact of the imputation on the composite indicator results. To discuss the presence of outliers in the dataset.
4. Multivariate analysis Should be used to study the overall structure of the data set, assess its suitability, and guide subsequent methodological choices (e.g., weighting, aggregation).	To check the underlying structure of the data along the two main dimensions, namely individual indicators and countries (by means of suitable multivariate methods, e.g., principal components analysis, cluster analysis).

	To identify groups of indicators or groups of countries that are statistically "similar" and provide an interpretation of the results. To compare the statistically-determined structure of the data set to the theoretical framework and discuss possible differences.
5. Normalisation Should be carried out to render the variables comparable.	To select suitable normalisation procedure(s) that respect both the theoretical framework and the data properties. To discuss the presence of outliers in the dataset as they may become unintended benchmarks. To make scale adjustments, if necessary. To transform highly skewed indicators, if necessary.
6. Weighting and aggregation Should be done along the lines of the underlying theoretical framework.	To select appropriate weighting and aggregation procedure(s) that respect both the theoretical framework and the data properties. To discuss whether correlation issues among indicators should be accounted for. To discuss whether compensability among indicators should be allowed.
7. Uncertainty and sensitivity analysis Should be undertaken to assess the robustness of the composite indicator in terms of the mechanism for including or excluding an indicator, the normalisation scheme, the imputation of missing data, the choice of weights, the aggregation method.	To consider a multi-modeling approach to build the composite indicator, and if available, alternative conceptual scenarios for the selection of the underlying indicators. To identify all possible sources of uncertainty in the development of the composite indicator and compute the composite scores and ranks with uncertainty bounds. To conduct sensitivity analysis of the inference (assumptions) and determine what sources of uncertainty are more influential in the scores and/ ranks.

8. Back to the data Is needed to reveal the main drivers for an overall good or bad performance. Transparency is primordial to good analysis and policymaking.	To profile country performance at the indicator level so as to reveal what is driving the composite indicator results. To check for correlation and causality (if possible). To identify if the composite indicator results are overly dominated by few indicators and to explain the relative importance of the sub-components of the composite indicator.
9. Links to other indicators Should be made to correlate the composite indicator (or its dimensions) with existing (simple or composite) indicators as well as to identify linkages through regressions.	To correlate the composite indicator with other relevant measures, taking into consideration the results of sensitivity analysis. To develop data-driven narratives based on the results.
10. Visualisation of the results Should receive proper attention, given that the visualisation can influence (or help to enhance) interpretability.	To identify a coherent set of presentational tools for the targeted audience. To select the visualisation technique which communicates the most information. To present the composite indicator results in a clear and accurate manner.

Source: OECD (2008), 'Handbook on Constructing Composite Indicators Methodology and User Guide'

Caution on choosing variable

1) Whenever we choose any variable or a particular dimension for the index, we have to justify the inclusion of the variable into the index. This justification should come from empirical evidence or policy based research studies or from theoretical explanations.

2) If the variable is not unidirectional, the entire variables used should be converted to unidirectional. For example, in construction of the food security index, two variables like 'proportion of agricultural worker to total workers' and per capita value of agricultural output' is used for index. Here the first variable is a negatively directional whereas the second variable is positively directional. In this case we have to convert the entire variables into either positive direction or negative direction. If we want to convert this to positive direction, the first variable which has a negative direction should be deducted from 100. On the other hand, if we want to convert the all the variables into negative direction, we have to work out the reciprocal of per capita value of agricultural output.

10.4 DEALING WITH MISSING VALUES AND OUTLIERS

After selecting the indicators, you have to have a clear idea on missing values for each selected variable. Data can be missing in a random or non random fashion. In such situation, easy solution is to drop cases for which data is missing. However, before dropping the variable, you have to see the number of cases for which values are missing, because exclusion of such households for which data is missing could significantly lower sample size. Further, deleting such household may lead to some biases also. For example, in livelihood study of households, if there is a missing value, we have to find out the frequency of missing value because in many cases, the chances of missing value is high for lower economic class as compared to upper economic class. In such cases if we delete the households having missing value, it may result biases towards upper economic classes of household. In case, inclusion or exclusion of households having missing value has a little or no impact on the final result, we can delete such cases.

The second solution is to impute the missing value by applying some methodology or logic. But use of imputation is more common in academic indices and data sets. Here you can try to impute the missing value which should be as accurate as possible..Consider calculation of the consumption expenditure of 50 households of which about 5 cases are missing. In this case in imputing we can substitute the average value of consumption expenditure and replace the missing value by the averages. But for a more accurate estimation we can group theses 50 households on the basis of value of assets holding and see to which asset category these 'MPCE missing' households belonged to. Then we can substitute the average consumption expenditure of the respected asset group.

Such approach has been used in a number of more recent indices such as the Corruption Perceptions Index (CPI) developed by Transparency International and the World Bank's Worldwide Governance Indicators (WGI). Such approaches carry a dual advantage. They allow scores to be estimated for a maximal number of countries, and can use a broader range of indicators to triangulate indices for nebulous constructs.

Many a time, outliers may also disturb the analysis. For example, if the income of four persons is 18000, 17000, 18500 and 19000 then the average is 18125. If we include the income of fifth person as 60000 then the average turns out to be 26500. In such matters, you have to drop the cases with high extreme value. Both in case of missing value and outlier you are expected to document and explain the selected imputation procedures and the results in detail.

Check Your Progress 1

1) What do you understand by the term Composite Index?

……..

……..

2) Give some examples of Composite Indices.

……

……

……

3) List the steps involved in construction of a Composite Index.

……

……

……

4) State the alternative methods to deal with the missing values of the selected variables.

……

……

……

10.5 METHODS TO CONSTRUCT COMPOSITE INDEX

The various methods are used in construction of composite indices. These include: Simple Ranking Method, Indices Method, Mean standardisation method, Range Equalisation Method, and Principal Component Analysis Method. Let us discuss one by one with simple examples.

10.5.1 Simple Ranking Method

Rank Method is one of the simplest methods to analyse the status of a region/district/state/country. In this method, the first step is to arrange/convert the variable into unidirectional for each district or state. Let us take an example of development status of different districts in Odisha state. First, we have to convert the relevant variables in either positive or negative direction. For example, the variable says 'the proportion of agricultural labour is negatively associated with the development of a region. Hence we have to convert this variable to positive direction by deducting this variable from 100 and change the labelling of variable as 'proportion of other than agricultural labour to total labour'. Higher the value of this variable, higher the development of the district or region. Depending on the value, each variable was ranked in the similar manner. Highest rank (1st) was given to the variables with high value and vice-à-versa. Alternatively, we can do reverse ranking of the variable i.e. highest rank is given to the variables with lowest value. The individual ranks are added to get the total rank value for the district. This has been illustrated in the example given in Table 10.1. Our objective is to find out most backward tribal district in Odisha (having more than 50% of tribal population to total population). A total of 11 districts qualify for tribal dominant area having 50% or more proportion of population. We have selected 10 variables for identifying most backward tribal areas. Each of the variables for 11 tribal

dominated districts was ranked according to total value of variable. The individual rank of all the variables of all the districts is given in Table 10.2. As all the variable are unidirectional, rank 1 is given to district having higher value and vice-a-versa. After doing so we can add together the values of all the variables and find out the average rank for each district (Sum of total rank divided by number of variable i.e. 10). The average value of variable shows the status of district on development indicators.

Table 10.1: District-wise Development Indicators of Tribal Concentrated Districts in Odisha

District	% of other than agricultural labour to all labour	% of net irrigated area to total sown area	Per capita value of agricultural output	Per capita consumption expenditure	Female Literacy rate	Women work force participation rate	% of household getting safe drinking water	Percentage of villages having access to paved road	Percentage of villages having access to primary health centre.	Average casual wage rate
Actual Value										
Malkangiri	73	31	1304	201	18	67	82	21	16	36
Rayagada	50	23	479	201	18	67	78	34	13	37
Sundargarh	61	19	537	280	43	58	57	38	28	32
Nabarangapur	46	6	839	201	18	66	80	45	28	37
Kandhamal	62	14	410	201	33	68	32	21	19	36
Koraput	55	31	611	201	16	67	67	22	22	40
Mayurbhanj	60	30	572	280	35	62	44	42	21	31
Gajapati	52	25	529	331	24	77	43	29	20	34
Sambalpur	62	34	1075	280	50	62	56	31	13	39
Kendujhar	60	23	537	280	44	46	52	46	15	34
Jharsuguda	68	20	662	280	54	43	63	47	32	39
Average	59	23	687	249	32	62	60	34	21	36

Table 10.2: District-wise Development Scenario of Tribal Odisha by Simple Ranking Method

District	% of other than agricultural labour to all labour	% of net irrigated area to total sown area	Per capita value of agricultural output	Per capita consumption expenditure	Female Literacy rate	Women work force participation rate	% of household getting safe drinking water	Percentage of villages having access to paved road	Percentage of villages having access to primary health centre	Average casual wage rate	Average Rank/indices
Malkangiri	1	2	1	7	8	5	1	10	8	7	5.0
Rayagada	10	7	10	7	9	3	3	6	11	4	7.0
Sundargarh	5	9	7	2	4	9	6	5	3	10	6.0
Nabarangapur	11	11	3	7	10	6	2	3	2	5	6.0
Kandhamal	3	10	11	7	6	2	11	11	7	6	7.4
Koraput	8	3	5	7	11	4	4	9	4	1	5.6
Mayurbhanj	7	4	6	2	5	7	9	4	5	11	6.0
Gajapati	9	5	9	1	7	1	10	8	6	8	6.4
Sambalpur	4	1	2	2	2	8	7	7	10	3	4.6
Kendujhar	6	6	8	2	3	10	8	2	9	9	6.3
Jharsuguda	2	8	4	2	1	11	5	1	1	2	3.7

By this method, the most backward tribal district is Jharsuguda and most developed district is kandhamal.

10.5.2 Indices Method

The indices method is another simple method for calculating the status of development of an area/district or state. Like rank method, here also, at the first stage the variables are converted into one direction. After converting the variable into one (positive or negative) direction, we calculate the index. In this method we have to convert the district figures basing the average figure for the entire 11 districts as 100. Let us take an example of Table 10.1. If the proportion of other than agricultural labour to total labour in malkangiri district is 73 and the average of that variable for all the 11 districts is 59, then the indices of that variable for Malkangiri district will be:

$$\frac{100}{59} \times 73 = 124.$$

In the same manner, we can find out the indices for all the variables for 11 districts as is given in Table 10.3. The final index of development for the districts can be obtained by taking the arithmetic mean for all the indicators (given in last column of Table 10.3). After working out the average value for all the districts, we can rank all the districts. The district with highest index

will be most developed and the district with lowest index will be least developed.

Table 10.3: District-wise Development Scenario of Tribal Odisha by Indices Method

District	% of other than agricultural labour to all labour	% of net irrigated area to total sown area	Per capita value of agricultural output	Per capita consumption expenditure	Female Literacy rate	Women work force participation rate	% of households getting safe drinking water	Percentage of villages having access to paved road	Percentage of villages having access to primary health centre	Average casual wage rate	Average Rank/indices
Malkangiri	124	133	190	81	57	108	138	62	78	99	107.0
Rayagada	85	99	70	81	57	108	131	98	64	103	89.6
Sundargarh	104	83	78	113	135	93	96	111	135	90	103.6
Nabarangapur	78	25	122	81	56	107	135	132	137	102	97.4
Kandhamal	106	59	60	81	102	109	54	60	92	100	82.3
Koraput	93	133	89	81	49	108	113	65	108	113	95.1
Mayurbhanj	101	129	83	113	109	101	74	123	103	86	102.2
Gajapati	89	106	77	133	76	123	73	86	95	95	95.3
Sambalpur	105	147	157	113	155	100	95	90	65	108	113.3
Kendujhar	102	99	78	113	136	74	88	136	71	94	99.1
Jharsuguda	114	88	96	113	167	69	105	138	153	109	115.3

In our example Kandhamal is the most backward with index (82.3) and Jharsuguda (115.3) is the most developed district.

Let us make a comparison between the results obtained by rank method and indices method. Results in Table 10.4 shows that Jharsuguda, Sambalpur, and Malkangiri are most developed in both the methods whereas Raigada and Kandhamal are most backward in both the methods. While comparing both rank and indices method, we can also run correlation between the two set of indices. If the correlation is high then we can say that the finding in both methods is almost similar.

Table 10.4: A Comparison of Simple Ranking Method and Indices Method

District	Status of District in Rank Method	District	Status of District in Indices Method
Jharsuguda	3.7	Jharsuguda	115
Sambalpur	4.6	Sambalpur	113

Malkangiri	5.0	Malkangiri	107
Koraput	5.6	Sundargarh	104
Sundargarh	6.0	Mayurbhanj	102
Nabarangapur	6.0	Kendujhar	99
Mayurbhanj	6.0	Nabarangapur	97
Kendujhar	6.3	Gajapati	95
Gajapati	6.4	Koraput	95
Rayagada	7.0	Rayagada	90
Kandhamal	7.4	Kandhamal	82

Note: Correlation of status of district in both rank and index method is 0.927

10.5.3 Mean Standardisation Method

Mean standardisation method is another simple method used both as a process of normalisation and also a composite index. In this method we normalise the value of each variable and than work out the average of the normalised value for all the variables. The average of normalised values will be the composite index value. The normalisation is done by dividing the actual value of variables by their respective means. Let us take an example from Table 10.1. For example, the standardised value of the variable '% of other then agricultural labour to all workers' for Malkangiri district can be found out by

Normalised value MS Method=Actual value (73)/Mean value (59) =1.237

In the same process we can find out the normalised value of the entire variable for 11 selected districts. The standardised value is given in table 10.5. After finding out the normalised value of all the variables we can take an average of all the variables for all the districts (given in last column of Table 10.5). This will be the composite index value.

Table 10.5: Index Value in Mean Standardisation Method

District	Percentage of other than agricultural labour to all labour	Percentage of net irrigated area to total sown area	Per capita value of agricultural output	Per capita consumption expenditure	Female Literacy rate	Women work force participation rate	Percentage of household getting safe drinking water	Percentage of villages having access to paved road	Percentage of villages having access to primary health centre	Average casual wage rate	Composite Index
Malkangiri	1.237	1.332	1.899	0.808	0.561	1.079	1.379	0.614	0.775	1.003	1.069
Rayagada	0.847	0.988	0.697	0.808	0.561	1.079	1.312	0.995	0.630	1.030	0.895

Sundargarh	1.034	0.816	0.782	1.126	1.340	0.934	0.959	1.112	1.357	0.891	1.035
Nabarangapur	0.780	0.258	1.222	0.808	0.561	1.063	1.346	1.316	1.357	1.030	0.974
Kandhamal	1.051	0.602	0.597	0.808	1.028	1.095	0.538	0.614	0.921	1.003	0.826
Koraput	0.932	1.332	0.890	0.808	0.499	1.079	1.127	0.644	1.066	1.114	0.949
Mayurbhanj	1.017	1.289	0.833	1.126	1.091	0.999	0.740	1.229	1.018	0.863	1.020
Gajapati	0.881	1.074	0.770	1.331	0.748	1.240	0.723	0.848	0.969	0.947	0.953
Sambalpur	1.051	1.461	1.565	1.126	1.558	0.999	0.942	0.907	0.630	1.086	1.132
Kendujhar	1.017	0.988	0.782	1.126	1.371	0.741	0.875	1.346	0.727	0.947	0.992
Jharsuguda	1.153	0.859	0.964	1.126	1.683	0.693	1.060	1.375	1.551	1.086	1.155

10.5.4 Range Equalisation Method

Range Equalisation (RE) method otherwise known as max-min approach is adopted by UNDP in computation of Human Development Index. Under this approach, an index is constructed for each variable by applying Range Equalisation formula derived by UNDP. This is worked out by subtracting an indicator's minimum value from each observation and then dividing it by its range.

In determining the goal post (minimum and maximum value) mainly two principles are adopted. The first one is taking the highest and lowest district value as the maximum and minimum value for the goal post. The second one is based on the value judgement of the researcher i.e. the likelihood of a particular variable to increase if favourable situation happens or decrease if unfavourable situation happens. The projected highest value for a favourable situation is the highest goal post and the reverse is the lowest goalpost.

We can explain this by way of an example. In case of irrigated area to cropped area, if the researcher presumes that in a period say coming 5 years if there will be some favourable situation in improving irrigation like government programme or good rainfall, then the highest goalpost will be 55% of irrigated area to cropped area. The 55% will be the highest goalpost. On the other hand, if for example due to drought in coming 5 years, if the Researcher presumes that the irrigation proportion will reduce by 2%, then 2% will be the lowest goalpost.

$$\text{RE Index} = \frac{X_i - Min\ X}{Max\ X - Min\ X}$$

where X_i Value of the variable, min X- Minimum value of X in the scaling, max X- Maximum value of X in the scaling. The RE index is also a normalisation technique. Without it, a composite index can be biased towards an indicator with very high range. For example, the per capita value of agricultural output which is measured in rupees ranges from Rs. 410 to Rs. 1304. On the other hand, the variable proportion of female literacy rate is measured in percentage. Different variables measured in different units

sometimes give upward or downward biases to ensure the index value biasfree. For this in RE method, we convert all the variables into equal scaling from 0 to 1.

In RE method, as a first step we undertake scaling exercise. In undertaking the scaling procedure, desirable norms are followed for each indicator. In some cases, scaling of indicator is self-selected, while in others, element of value judgement is involved. This scaling exercise is called goal post where we identify the maximum and minimum goal. The goalpost basically visualises the extent of minimum value and maximum value in the future time period (say next 5 years). The scaling norm that we have adopted is given in Table 10.6. In the table the third and fourth column shows the maximum and minimum district value of selected 10 indicators. The first and second column is the maximum and minimum goal post.

Table 10.6: Construction of Food Security Radar

Variable	Description of Variables	Goalposts		District Value	
		Minimum	Maximum	District Minimum	District Maximum
Oth_agl	Percentage of other than agricultural labour to all labour	30	85	46	73
Irr	Percentage of net irrigated area to net cropped area	2	55	6	34
pcvao	Per capita value of agricultural output	200	2500	410	1304
Mpcce_ia	Inequality adjusted per capita consumption expenditure	150	450	201	331
Lit_f	Female Literacy (adult) rate	10	70	16	54
Wfpr_f	Women work force participation rate	30	85	43	77
Hhsdw	Percentage of household having access to safe drinking water	20	90	32	82
Paved_r	Percentage of villages having access to paved road	10	60	21	47
v_phcs	Percentage of villages having access to Primary health centre	10	50	13	32
Wage_c	Average Casual wage rate	25	70	31	40

Based on the goalpost, we normalise the value of the variable for all the districts and all the indicators. In our example, the values have been presented in Table 10.7.

Table 10.7: Index Value in Range Equalization Method

District	Percentage of other than agricultural labour to all labour	Percentage of net irrigated land to net sown area	Per capita value of agricultural output	Per capita consumption expenditure	Female Literacy rate	Women work force participation rate	Percentage of households getting safe drinking water	Percentage of villages having access to paved road	Percentage of villages having access to primary health centre	Average casual wage rate	Composite Index
Malkangiri	0.782	0.547	0.480	0.170	0.133	0.673	0.886	0.220	0.150	0.244	0.429
Rayagada	0.364	0.396	0.121	0.170	0.133	0.673	0.829	0.480	0.075	0.267	0.351
Sundargarh	0.564	0.321	0.147	0.433	0.550	0.509	0.529	0.560	0.450	0.156	0.422
Nabarangapur	0.291	0.075	0.278	0.170	0.133	0.655	0.857	0.700	0.450	0.267	0.388
Kandhamal	0.582	0.226	0.091	0.170	0.383	0.691	0.171	0.220	0.225	0.244	0.300
Koraput	0.455	0.547	0.179	0.170	0.100	0.673	0.671	0.240	0.300	0.333	0.367
Mayurbhanj	0.545	0.528	0.162	0.433	0.417	0.582	0.343	0.640	0.275	0.133	0.406
Gajapati	0.400	0.434	0.143	0.603	0.233	0.855	0.329	0.380	0.250	0.200	0.383
Sambalpur	0.582	0.604	0.380	0.433	0.667	0.582	0.514	0.420	0.075	0.311	0.457
Kendujhar	0.545	0.396	0.147	0.433	0.567	0.291	0.457	0.720	0.125	0.200	0.388
Jharsuguda	0.691	0.340	0.201	0.433	0.733	0.236	0.614	0.740	0.550	0.311	0.485

After calculating the index of each variable, we have averaged them to give each of the five dimensions of food security. The composite food security index is again derived by averaging the five dimensions.

The normalised value for the variable percentage of other than agricultural labour to total worker' for Malkangiri district is found out in the following manner.

$$\frac{(Actual\ value\ (73) - Minimum\ goalpost\ (30))}{(Maximum\ goalpost\ (85) - Minimum\ goalpost\ (30))} = 0.782$$

Likewise we have converted and worked out the normalised value of all the variables and for all the districts shown in Table 10.7.

Comparing MS and RE

The composite index value worked out by both the RE and MS methods can be compared and analysed (Table 12.8). We can also examine correlation between two indices. Here the correlation between final index value worked out by both the methods is very high (0.991).

Table 10.8: Comparison between RE and MS Methods

District	Composite Index Value by MS Method	Composite Index Value by RE Method
Jharsuguda	1.155	0.485
Sambalpur	1.132	0.457
Malkangiri	1.069	0.429
Sundargarh	1.035	0.422
Mayurbhanj	1.020	0.406
Kendujhar	0.992	0.388
Nabarangapur	0.974	0.388
Gajapati	0.953	0.383
Koraput	0.949	0.367
Rayagada	0.895	0.351
Kandhamal	0.826	0.300

Note: Correlation of index value is 0.991

The high degree of the correlation between the values of composite index computed by applying both methods separately indicate the high degree of homogeneity between the two approaches. However, the Range equalisation method is preferred because it accounts for wider variations and strong correlations to the PCA composite (PCA is a method which is far beyond the scope of present unit).

One important point should be kept in mind that the index value as discussed above are based on equal weights. We can also give weights to variables/components discussed in Section 10.5.

Check Your Progress 2

1) State the procedure to workout Composite Index by way of Simple Ranking Method.

   ……

   ……

2) How will you compute the final value of Composite Index by Indices Method?

   ……

   ……

   ……

3) What is the Distinction between Range Equalisation Method and Mean Standardisation Method?

   ……

   ……

10.6 ILLUSTRATIONS OF COMPOSITE INDICES

10.6.1 Physical Quality of Life Index (PQLI)

Physical Quality of Life Index (PQLI) was developed by famous economist Morris David in 1979 for 23 developed and developing countries. Morris David used the following three indicators to prepare a composite index known as Physical Quality of Life Index:

1) Life Expectancy Rate (LEI): Life expectancy means average number of year a person is expected to live. As per census of 2011, it is 66.8 years in India.

2) Infant Mortality Rate (IMI): It refers to the number of infants dying within one year of their birth out of every 1000 births. As per census report of 2011, it is 47 per 1000. Higher infant mortality is harmful for economic development.

3) Basic Literacy Rate (BLI):Any person above the age of 7 year who can read and write in any one language with an ability to understand, it is considered as literate. As per census 2011, it is 74.04% in India.

For each of the above indicator, the performance of individual country is rated on a scale of 1 to 100 where 1 represents the worst performance and 100 represent the best performance. PQLI is then constructed by averaging these three indicators giving equal weight to each of them. Morris David has given following formula to obtain PQLI

$$\text{PQLI} = \frac{LEI+IMI+BLI}{3}$$

Suppose, the LEI for India is rated as 75, the IMI is rated as 40 and BLI is rated as 65, then the PQLI will be 60 [(75+40+65)/3].

10.6.2 Human Development Index (HDI)

According to UNDP Report, the Human Development Index (HDI) is a summary measure of achievements in three key dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. The HDI is the geometric mean of normalised indices for each of the three dimensions. Steps involved in to calculating the HDI are:

Step 1: Creating the dimension indices

Minimum and maximum values (goalposts) are set in order to transform the indicators expressed in different units into indices on a scale of 0 to 1, as indicated below:

Dimension	Indicators	Maximum Value	Minimum Value
Health	Life expectancy at birth (years)	85	20

Education	Expected Years of Schooling	18	0
	Minimum Years of Schooling	15	0
Standard of Living	GDP per capita (PPPUS$)	75000	100

Having defined the minimum and maximum values, the dimension indices are calculated as:

Dimension index = $\frac{Actual\ value - Minimum\ value}{Maximum\ value - minimum\ value}$.

Step 2: Aggregating the dimensional indices to produce the Human Development Index

The HDI is the geometric mean of the three dimension indices

HDI = (*Health indicator* × *Education indicator* × *Income indicator*) $^{1/3}$

10.6.3 Gender Development Index (GDI)

The Gender Development Index (GDI) measures gender inequalities in achievement in three basic dimensions of human development: health, measured by female and male life expectancy at birth; education, measured by female and male expected years of schooling for children and female and male mean years of schooling for adults ages 25 years and older; and command over economic resources, measured by female and male estimated earned income.

Step 1: Estimating the female and male earned incomes

To calculate estimated earned incomes, the share of the wage bill is calculated for each gender. The female share of the wage bill (S_f) is calculated as follows:

$$S_f = \frac{W_f/W_m.EA_f}{W_f/W_m.EA_f + EA_m} = \frac{(W_f.EA_f)/W_m}{(W_f.EA_f + W_m.EA_m)/W_m} = \frac{Wage\ Income\ F}{Total\ Wage\ income}$$

Where W_f/W_m is the ratio of female to male wage, EA_f is the female share of the economically active population and EA_m is the male share of the economically active population. The male share of the wage bill is calculated as:

$$S_m = 1 - S_f$$

Estimated female earned income per capita (GNI_{pcf}) is obtained from GNI per capita (GNI_{pc}), first by multiplying it by the female share of the wage bill, S_f , and then rescaling it by the female share of the population, $P_f = N_f /N$, where N_f is number of females in total population N:

$$GNI_{pc_f} = GNI_{pc}.S_f/P_f$$

Estimated male earned income per capita (GNI_{pcm}) is obtained in the same way:

$$GNI_{pcm} = GNI_{pc}.S_m/P_m$$

Where, $P_m = N_m/N$; N_m being the number of males in total population.

Step 2: Normalising the indicators

To construct the female and male HDI values, first the indicators, which are in different units, are transformed into indices and then dimension indices for each sex, are aggregated by taking the geometric mean. The indicators are transformed into indices on a scale of 0 to 1 using the same goalposts that are used for the HDI, except life expectancy at birth, which is adjusted for the average five-year biological advantage that women have over men.

Dimension	Indicators	Maximum Value	Minimum Value
Health	Life expectancy at birth (years)	85	20
	Female	87.5	22.5
	Male	82.5	17.5
Education	Expected Years of Schooling	18	0
	Minimum Years of Schooling	15	0
Standard of Living	GDP per capita (PPPUS$)	75000	100

Having defined the minimum and maximum values, the dimension indices are calculated as:

$$\text{Dimension index} = \frac{Actual\ value - Minimum\ value}{Maximum\ value - minimum\ value}.$$

Step 3: Calculating the female and male Human Development Index values

The female and male HDI values (HDI_f and HDI_m) are the geometric means of the three dimensional indices for each gender

Step 4: Calculating the Gender Development Index

The GDI is simply the ratio of female HDI to male HDI:

$$GDI = \frac{HDI_f}{HDI_m}$$

10.7 MERITS AND LIMITATION OF COMPOSITE INDEX

One of the important merits of a composite index is that this can summarise the complex and multidimensional indicators into one indicator which helps the policy makers for implementation of a particular programme or policy decision. A composite index can easily be taken to interpret about the development or backwardness of a particular region or area or sector. The progress of a state in India say, e.g. Bihar can easily be accessed and we can compare the status of Bihar at two or more points of time. The composite index can reduce the number of indicators without reducing the underlying information base. The composite index fascilates communication to general public and promote accountability.

The composite index is not free from limitations. It has the risk of misleading policy message if it is poorly constructed or misinterpreted. The calculation may be misleading and faulty if sound statistical or conceptual principle is not applied. The assignment of weight on the basis of value judgement many a time creates a debate. This has always been a source of criticism. Again allocation of upper and lower bound (mainly in range equalisation method) is a point of criticism.

Check Your Progress 3

1) What is PQLI? Do you agree with the view that PQLI is a useful indicator of human development?

   ……

   ……

2) In construction of Human Development Index, how weights are assigned to the different indicators?

   ……

   ……

3) How the GDI measure is calculated?

   ……

   ……

4) State the limitations of Composite Index.

   ……

   ……

10.8 LET US SUM UP

Composite index is an important analytical technique to analyse the developmental related issues like status of development, food security, human development etc. at the district/region/state level. It is an expression of single score made of different scores measuring the various dimensions and indicators of a particular issue. This unit describes the process to derive composite index by a applying the various methods. These methods include simple ranking method, indices method, mean standardisation method, and range equalisation method. The main advantage of the rank and indices method is that they are easy to understand and interpret. The RE, MS and PCA methods use all the variables in reducing the dimensionality of the data. As the PCA method is beyond the scope of present unit, we have not illustrated its use here. These methods are very much useful in comparing across districts, states, countries or areas such as rural and urban. Again these methods are useful to compare over time by constructing composite index at two points of time by taking same indicators/variables.

10.9 EXERCISES

1) Think and collect some gender related variables from the data sources like women literacy, women workforce participation rate in the age group 15-59, mean year schooling, mortality rate, immunisation rate etc and find out the gender development index.
2) By using range equalisation method find out the food security index from the report prepared by IHD-WFP given in website

   'http://122.180.7.122/displaymorePub.asp?itemid=84&subchkey=11&chname=Publications.'

10.10 REFERENCES

1) Saltelli, G. Munda and M. Nardo (2006). 'From Complexity to Multidimensionality: the Role of Composite Indicators for Advocacy of EU Reform', Tijdschrift voor Economie en Management Vol. LI, 3.
2) Downloaded from http://www.econ.kuleuven.be/rebel//jaargangen/2001-2010/2006/TEM%202006-3/TEM_2006-3_03_Saltelli.pdf.
3) Planning Commission (2001), 'National Human Development Report', Government of India downloaded from

   http://planningcommission.nic.in/reports/genrep/index.php?repts=nhdcont.htm
4) OECD (2008), 'Handbook on Constructing Composite Indicators Methodology and User Guide', Organization For Economic Co-operation and Development.

10.11 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) See Section 10.2
2) See Section 10.2
3) See Section 10.3
4) See Section 10.4

Check Your Progress 2

1) See Sub-section 10.5.1
2) See Sub-section 10.5.2
3) See Sub-section 10.5.4

Check Your Progress 3

1) See Sub-section 10.6.1
2) See Sub-section 10.6.2
3) See Sub-section 10.6.3
4) See Section 10.7

UNIT 11 ANALYSIS OF QUALITATIVE DATA

Structure

11.0 Objectives

11.1 Introduction

11.2 Qualitative Research

- 11.2.1 What is Qualitative Research?
- 11.2.2 Purpose of Qualitative Research
- 11.2.3 Advantages and Disadvantages of Qualitative Research
- 11.2.4 Qualitative vs. Quantitative Research

11.3 Qualitative Data: Research Methods

- 11.3.1 Phenomenology
- 11.3.2 Ethnography
- 11.3.3 Grounded Theory
- 11.3.4 Participatory Action Research(PAR)
- 11.3.5 Case Studies

11.4 Qualitative Data and Techniques

- 11.4.1 Qualitative Data
- 11.4.2 Types of Qualitative Data
- 11.4.3 Qualitative Sampling Techniques

11.5 Qualitative Data Collection Methods

- 11.5.1 Observations
- 11.5.2 Interviews
- 11.5.3 Focus Group Discussions
- 11.5.4 Written Documents
- 11.5.5 Other Methods

11.6 Qualitative Data Analysis: Approaches and Techniques

- 11.6.1 Data Analysis Approaches
- 11.6.2 Data Analysis Techniques
  - 11.6.2.1 Content Analysis
  - 11.6.2.2 Narrative Analysis
  - 11.6.2.3 Discourse Analysis
  - 11.6.2.4 Framework Analysis
  - 11.6.2.5 Grounded Theory Analysis
  - 11.6.2.6 Thematic Analysis

11.7 Qualitative Data Analysis: Procedure and Computer Softwares

- 11.7.1 Data Analysis Procedure
- 11.7.2 Types of Software Used for Qualitative Analysis

11.8 References

11.9 Answers or Hints to Check Your Progress Exercises

11.0 OBJECTIVES

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

- explain qualitative research design, its importance and purpose;
- distinguish between qualitative and quantitative research designs;
- understand the qualitative data, its type and sampling procedures;
- get an idea about various kinds of qualitative research designs one can adopt;
- discuss methods of collecting qualitative data; and
- learn various steps and techniques of Qualitative Data Analysis and available software one can use.

11.1 INTRODUCTION

Research is a systematic inquiry used to describe, explain, predict or control some observed phenomenon. In any research study, generally researchers begin by selecting a broad research topic and conducting a literature review to build up the researcher's knowledge base and to ensure the significance of the research. Thereafter, the researcher develops a research problem related to the topic and create specific questions. The next step is to develop the research design. Now, here the most important question that arises is, which research design to be selected that could be most suitable to address the particular research problem? The choice need to be made between quantitative, qualitative or mixed methods. However, before making choices, it is very important to have clear understanding about each of these methods separately. Quantitative methods have already been explained in Unit 2, Block 2 and Block 3. This Unit introduces Qualitative research designs. The questions like: What do we mean by Qualitative research design and qualitative data? How does it differ from Quantitative techniques? Why this research design is important/required? What are types of qualitative data and how one can collect this type of data? What are its sources, etc. will be explained in this Unit. Further to this, within Qualitative method, there are several research designs that may be considered. We shall cover various approaches of qualitative research design, data analysis techniques and procedure in this Unit. Let us begin with explaining the concept of Qualitative research design.

11.2 QUALITATIVE RESEARCH

11.2.1 What is Qualitative Research?

According to Pope and Mays (1995), Qualitative research is defined as "Development of concepts which help us to understand social phenomena in natural (rather than experimental) settings, giving due emphasis to the

meanings, experiences and views of the participants". The qualitative research focuses on the "why" rather than the "what" of social phenomena and relies on the direct experiences of human beings in their everyday lives. The common assumption of this approach is that knowledge is subjective rather than objective and the researcher learns from the participants in order to understand the meaning of their lives.

This research design generates data about human groups in social settings. Hence instead of applying logical and statistical procedures, multiple systems of inquiry including biography, case study, historical analysis, discourse analysis, ethnography, grounded theory and phenomenology are used. We shall discuss these methods later in this Unit.

11.2.2 Purpose of Qualitative Research

Qualitative research methods are a common choice in social science research problems and are often used to study ideas, beliefs, human behaviours and other research questions that do not involve studying the relationship between variables. These methods are chosen when

- The goal of the research problem is to examine, understand and describe a phenomenon.
- To generate a sample which allows understanding the social process of interest.
- Qualitative research is aimed at gaining deep understanding of a specific organisation or event, rather than surface description of a large sample of a population.
- It aims to provide an explicit rendering of the structure, order, and broad patterns found among a group of participants.
- Qualitative research enables us to make sense of reality, to describe and explain the social world and to develop explanatory models and theories.
- It is the primary means by which the theoretical foundations of social sciences may be constructed or re-examined.

11.2.3 Advantages and Disadvantages of Qualitative Research

Research is the means by which discoveries are made, ideas are confirmed or refuted, events controlled or predicted and theory developed or refined, all of which contributes to the development of knowledge. However, no single research approach fulfills all of these functions as each of them have certain pros and cons. The advantages and disadvantages of qualitative research are given in this section.

Advantages

- Through qualitative research, it becomes possible to understand attitudes.
- Content Creator: This approach allows for genuine ideas to be collected from specific socio-economic demographics that can further be used to create valuable content.

- Saves money: The qualitative research process uses a smaller sample size than other research methods, thereby leading to lower research costs.
- It can provide deeper insights related to research topic.
- Due to open-ended questions, the qualitative research process allows researchers to mine the authentic and emotional data that drives decision-making responses.
- It has flexibility: There isn't a rigid structure to the qualitative research process. Because of this flexibility, trained researchers are permitted to follow-up on any answer they wish to generate more depth and complexity to the data being collected.
- It offers predictive qualities: The data which is gathered through qualitative research is perspective-based, which is why it has a predictive quality to it.
- It allows for human instinct to play a role: The qualitative data often provides higher level of accuracy and authenticity than any other form of data offered.
- It can be based on available data, incoming data, or other data formats: The qualitative research method does not require a specific pattern or format for data collection. If researchers feel like they are not generating useful results from their efforts, they can change their processes immediately. There are more opportunities to gather new data when using this approach.

Disadvantages:

- It can be time consuming: Because researchers follow numerous tangents when collecting data, it takes more time to gather it. Sorting through all of that extra data takes time.
- Masses of data has to be transcribed which is tedious and time consuming process.
- Comparatively, data coding is a difficult task in this approach.
- This approach may not be applicable to widely dispersed social settings.
- Usually this gives only nominal level data, difficult to quantify and control researcher bias.
- The qualitative research process does not provide statistical representation. It will only provide research data from perspectives only. Responses with this form of research cannot usually be measured.
- In case researchers are unable to see necessary data when they observe it, there is risk of losing important observation which lessens the accuracy of the results from the qualitative research efforts. That could even lead some research efforts toward false conclusions.
- This approach might not be effective where important decision has to be made, for which numerous perspectives are often required.

- Difficult to replicate results: Because qualitative research is based on individual perspectives, it is almost impossible to duplicate the results that are found. Even the same person may have a different perspective tomorrow than they had today. That means the data collected through qualitative research can be difficult to verify, which can lead some to question the conclusions that researchers generate through this process.
- It can create misleading conclusions: Since qualitative research constitutes small sample so there is always a chance that the research sample may include like-minded people while completely ignore the perspectives of those on the other side. There is no absolute way to know if the conclusions generated through qualitative research can apply to an entire demographic region.
- Research conclusions may be influenced by researcher's bias.
- Risk of non-acceptability: Even though there is a certain authenticity to qualitative research, there is also a certain subjectivity to it. Because of this nature, the data collected may not be accepted. If similar qualitative research efforts do not produce similar results, the data originally collected might even be rejected.

11.2.4 Qualitative vs. Quantitative Research

For people who are new to research methods, an understanding of the difference between qualitative and quantitative data analysis is very important. The simple distinction between these two methods is that qualitative data analysis deals with analysis of subjective and non-numerical data, while quantitative data analysis focuses on the measurement and analyses of data through a numerical or statistical means. Each of the qualitative and quantitative paradigms has its own set of assumptions, established methodologies and set of experts, which is very important to understand by the researcher. Some of the distinctions between both of these paradigms are discussed in this section.

	Quantitative Research	Qualitative Research
Scientific Method	Deductive 'top-down'	Inductive 'bottom-up'
Research Objective	Description, prediction, explanation	Description, exploration and discovery
Focus	Narrow angle lens, hypothesis testing	Wide and deep angle, examining phenomena
Purpose	To examine the relationship between variables with the primary goal to analyse and represent the relationship mathematically through statistical analysis.	Primarily concerned with the development of description of an observed phenomenon to generate solid theory as an outcome, or the product of their research.

Requirement	Requires the use of standardised measures in order to fit the different perspectives and experiences of people into a limited number of predetermined response categories to which numbers are assigned.	Requires the researcher to study selected issues in depth and detail without being constrained by pre-determined categories of analysis.
Emphasis	Emphasis is on testing of the theory.	Emphasis is on the construction of the theory.
Theory validation	Here, pertinent knowledge from previous research and from everyday life is organised into theory — given what is known and one's best guess about reality — to build a cogent and best argument that may answer the question. This framework is then tentatively constructed as conjecture, systematically tested and subsequently revised in light of these experimental results.	Here, theory is constructed systematically, and even 'tested' or verified during the process of construction. Qualitatively derived theory has minimal risk of invalidity.
Nature of Observation	Study behaviour under controlled conditions	Study context in which behaviour occurs in natural setting thereby valuing openness and flexibility.
Process	The research methods are set before observation begins and specify the methods of observation which may be used and the type of data which may be collected. Observations are collected before analysis begins. After analysis is complete, no more observations are taken.	The research methods suggest the type of methods of observation that may be used and the type of data which may be collected. Analysis begins as soon as data begin to be collected. Analysis and data collection proceed in a cyclical fashion, where preliminary analysis informs subsequent data collection and so forth.

Data Collection Form	Structured and validated data collection instruments	Unstructured or semi structured In-depth interviews, observation, field notes, FGDs
Data Type	Variable	Words, images, categories
Data Generalisation	The statistical data of quantitative methods obtained from a great many people results in a broad, generalisable set of findings that are concise and brief.	It produces a large amount of detailed information about a smaller number of people that results in rich understanding but reduces generalisability.
Data Analysis and Results	Identify statistical relationship	Particularistic, viewpoint representation
Report Form	Statistical	Narrative, contextual

11.3 QUALITATIVE DATA: RESEARCH METHODS

Qualitative research designs provides general way of thinking about conducting qualitative research. In this section, five major qualitative research designs are introduced.

11.3.1 Phenomenology

Phenomenology means study of individual's life experiences of events. Phenomenology is a school of thought that emphasises a focus on people's subjective experiences and interpretations of the world. Under this approach researcher studies a human experience at an experiential level such as understanding what it means for a woman to lose a child or experience of AIDS care. It is about understanding the essence or meaning of the experience. Phenomenology is sometimes considered a philosophical perspective as well as an approach to qualitative methodology. It has a long history in several social research disciplines including psychology, sociology and social work. A phenomenological study attempts to set aside biases and preconceived assumptions about human experiences, feelings and responses to particular situation.

11.3.2 Ethnography

The ethnographic approach to qualitative research comes largely from the field of anthropology and means 'portrait of a people'. Ethnography is a methodology for descriptive studies that looks at the people, cultures and commonalities of shared experiences. Originally, the idea of a culture was tied to the notion of ethnicity and geographic location, but it has been broadened to include virtually any group or organisation. That is, we can study the "culture" of a business or defined group. Ethnography is an extremely broad area with a great variety of practitioners and methods. However, the most common ethnographic approach is participant observation as a part of field research. The ethnographer becomes immersed in the culture as an active participant and records extensive field

notes, therefore, ethnography is very time consuming process. The risk of using ethnographic research is that the researcher may not fully understand or be familiar with the social norms of those they are researching and therefore there is risk of misinterpretation.

11.3.3 Grounded Theory

Under this approach, the focus is on to develop an understanding of a phenomenon or situation so as to develop a theory/model for items such as factors, a form of interaction, or a process. Grounded theory is a qualitative research approach that was originally developed by Glaser and Strauss in the 1960s. Grounded theory is a complex *iterative* process. The research begins with the raising of *generative questions* which help to guide the research but are not intended to be either static or confining. As the researcher begins to gather data, *core theoretical concept(s)* are identified. Tentative *linkages* are developed between the theoretical core concepts and the data. This early phase of the research tends to be very open and can take months. Later on the researcher is more engaged in verification and summary.

There are several key analytic strategies under grounded theory:

- *Coding* is a process for both categorising qualitative data and for describing the implications and details of these categories. Initially one does *open coding*, i.e. developing some initial categories and later one moves to more *selective coding* where one systematically codes with respect to a core concept.
- *Memoing* is a process for recording the thoughts and ideas of the researcher as they evolve throughout the study. It is similar to preparing extensive marginal notes and comments.
- *Integrative diagrams and sessions* are used to make sense of the data with respect to the emerging theory. The diagrams can be any form of graphic that is useful at that point in theory development, for e.g. concept maps or directed graphs or even simple cartoons that can act as summarising devices. This integrative work is best done in group sessions where different members of the research team are able to interact and share ideas to increase insight.

11.3.4 Participatory Action Research (PAR)

Participatory action research (PAR) is an approach that emphasises collective inquiry and experimentation grounded in experience and social history. It seeks to understand the world by trying to change it, collaboratively and following reflection. It involves researchers and participants working together to understand a problematic situation and change it for the better. In other words, it involves participation and action. PAR focuses on social change that promotes democracy and challenges inequality; is context-specific, often targeted on the needs of a particular group; is an iterative cycle of research, action and reflection; and often seeks to 'liberate' participants to have a greater awareness of their situation

in order to take action. PAR promotes changes desired by the group in the form of policies, programmes or research developments.

11.3.5 Case Studies

Case study involves in-depth investigation of a single or small number of units at a point in time. It is an intensive study of a specific individual, situation or specific context. For example, evaluation of a unique teaching methodology in some region. Case studies are conducted to collect a variety of material in a specific and bounded time period. This is also used for historical studies, when collecting historical data to understand and learn from the past.In qualitative research, data is gathered by taking a close look and an in-depth analysis of a “case study” or “case studies” – the unit or units of research that may be an individual, a group of individuals, or an entire organisation. The researcher may employ combination of methods for analysing case studies such as interviewing, floating questionnaires, or conducting group discussions in order to gather data.

Check Your Progress 1

1) Explain the term ‘Qualitative research’ and how it differs from Quantitative research design?

.....

.....

.....

.....

2) What is the difference between phenomenology and ethnography?

.....

.....

.....

.....

.....

3) What do you mean by grounded theory: Discuss its various strategies and in what way ethnography is similar to grounded theory?

.....

.....

.....

.....

.....

.....

4) Is Participatory Action Research (PAR) approach qualitative in nature? Explain?

.....

.....

.....

5) What are the advantages and disadvantages of using case studies?

.....

.....

.....

11.4 QUALITATIVE DATA AND TECHNIQUES

The above sections have already discussed about qualitative research paradigm and various designs, researchers can adopt to conduct qualitative research studies. The next most important step is to have clarity on qualitative data, its types and sampling procedures. All of these are discussed here in this section in brief.

11.4.1 Qualitative Data

Qualitative data is also known as descriptive data. It is non-numerical in nature or in a form that cannot be easily reduced to numbers but can be observed or recorded. Qualitative data captures concepts, opinions, values and behaviours in the social context. Qualitative data is all about the emotions or perceptions of people that are documented in a form which can help the researchers identify and deal with the problem effectively and efficiently. Some examples of qualitative data include transcripts of individual interviews and focus groups, field notes from observation, document copies, audio/video recordings etc.

Qualitative data in statistics is also known as categorical data because it can be arranged categorically based on the attributes and properties of a thing or a phenomenon. For example, think of students who have been asked to provide their feedback on quality of teaching practices in schools. The student may give feedback based on teacher's knowledge, their attitudes, friendly nature, ability to make students understand and clarity in pronunciation, without giving grades to all teachers, this is considered as an example of qualitative data.

It's pretty easy to understand the difference between qualitative and quantitative data. Qualitative data does not include numbers in its definition of traits whereas quantitative data is all about numbers. An example of qualitative data is the taste of a cake that can be categorised as chocolate, pineapple, vanilla etc. Qualitative data is important in determining the particular frequency of traits or characteristics. It allows the statistician or the researchers to form parameters through which larger data sets can be

observed. Qualitative data provides the means by which observers can quantify the world around them.

11.4.2 Types of Qualitative Data

The type of qualitative data along with its sources is discussed in this section. Qualitative data can broadly be classified into following categories:

- **Structured text:** This kind of data is available in the form of certain writings, stories, comments, news articles, books, memos, diaries etc. Some of the sources of collecting this data are newspapers, libraries, magazines etc.
- **Unstructured text:** Unstructured text may be collected in the form of personal interviews conducted with the respondents, transcriptions or focus groups conversation. Researchers generally ask open ended questions here to explore the idea and does not include any standard set of questions.
- **Audio-video recordings:** Consist of images or sounds that researchers collect to help them understand the central phenomenon under the study. This includes graphics, art, pictures, visuals, films, videotapes etc.

11.4.3 Qualitative Sampling Techniques

There are number of sampling techniques available through which researchers can employ for collecting the data. These sampling techniques are discussed here in this section.

- **Snowball/chain sampling:** Snowball Sampling is also known as chain referral sampling. In this method, the participants refer the researcher to others who may be able to potentially contribute or participate in the study. This method often helps researchers find and recruit participants that may otherwise be hard to reach. For example, there might be in-depth interview of government school principals. So one school principal may refer to other principals in his/her circle.
- **Extreme/Deviant Case sampling:** It is the form of purposeful sampling in which researcher studies the outlier case or one that displays the extreme characteristics. It captures certain kind of crisis events for example survivors of an aeroplane crash or of a building collapse.
- **Convenience sampling:** Convenience sampling is a non-probability sampling technique where subjects are selected because of their convenient accessibility and proximity to the researcher. The subjects are selected just because they are easiest to recruit for the study and the researcher did not consider selecting subjects that are representative of the entire population. For example, using subjects that are selected from a clinic, a class or an institution that is easily accessible to the researcher.
- **Opportunistic sampling:** This is a kind of purposeful sampling which is undertaken after the research begins to take advantage of unfolding events that will help answer the research questions. For example, let's

say researcher wanted to study the learning behaviour of disabled children. Instead of taking a random sample of these children, researcher may decide to visit the nearby school of disabled children and select one whole class as his/her sample.

- **Homogeneous sampling:** Under this method, researcher purposefully samples individuals or sites based on membership in a subgroup that has defining characteristics/traits. For example, taking views of adult female cancer patients of given district/state.
- **Maximum variant sampling:** It is a purposeful sampling strategy in which the researcher samples cases or individuals that differ on some characteristics or traits. For example, based on different age-groups or socio-economic class or from different job profiles, sample cases are identified.

Check Your Progress 2

1) What do you mean by qualitative data? Explain with the help of an example.

.....

.....

.....

2) Discuss the types and sampling techniques of qualitative data?

.....

.....

.....

3) What should be the adequate sample size for conducting qualitative research?

.....

.....

.....

11.5 QUALITATIVE DATA COLLECTION METHODS

Qualitative research does not introduce treatments or manipulate variables, or impose the researcher's operational definitions of variables on the participants. Rather, it lets the meaning emerge from the participants. It is more flexible in that it can adjust to the setting. Concepts, data collection tools, and data collection methods can be adjusted as the research progresses. Qualitative research aims to get a better understanding through first-hand experience, truthful reporting, and quotations of actual conversations. It aims to understand how the participants derive meaning from their surroundings, and how their meaning influences their behaviour. Qualitative research uses various data collection method as discussed in this section.

11.5.1 Observations

Observation is the selection and recording of behaviours of people in their natural environment. Observation is useful for generating in-depth descriptions of organisations or events, for obtaining information that is otherwise inaccessible, and for conducting research when other methods are inadequate. Observation is used extensively in studies by psychologists, anthropologists, sociologists, and programme evaluators. Direct observation reduces distortion between the observer and what is observed that can be produced by an instrument (e.g., questionnaire). The context or background of behaviour is included in observations of both people and their environment. The researcher gets close enough to study subjects to observe (with/without participation) usually to understand whether people do what they say they do, and to access inferred knowledge of subject. Direct observation differs from interviewing in that the observer does not actively query the respondent in this case. The data can be recorded in many of the same ways as interviews (stenography, audio, video) and through pictures, photos or drawings (e.g., those courtroom drawings of witnesses are a form of direct observation). The advantage is that one can study actual behaviour and those individuals who have difficulty in verbalising ideas. The disadvantage is that it is limited to sites and situations where one can gain access.

11.5.2 Interviews

Interviews are one of the most commonly used data collection instrument for qualitative research, mainly because of its personal approach. This involves asking questions, listening to and recording answers from an individual or group on a structured, semi-structured or unstructured format in an in-depth generally in an informal manner. Unstructured interviewing involves direct interaction between the researcher and a respondent or group. The questions asked are mostly open-ended questions, spontaneous, with the interviewer letting the flow of the interview dictate the next questions to be asked.The purpose of the interview is to probe the ideas of the interviewees about the phenomenon of interest. Here, although the researcher may have some initial guiding questions or core concepts to ask about, there is no formal structured instrument or protocol. Moreover, the interviewer is free to move the conversation in any direction of interest that may come up. Unstructured interviewing is particularly useful for exploring a topic broadly. The only disadvantage is that because each interview tends to be unique with no predetermined set of questions asked of all respondents, it is usually more difficult to analyse unstructured interview data, especially when synthesising across respondents.

11.5.3 Focus Group Discussions

It involves focused (guided by a set of questions) and interactive session with a group of respondents, small enough for everyone to have chance to talk and large enough to provide diversity of opinion.The group is limited to 6-10 people and a moderator is assigned to moderate the ongoing discussion. Depending on the data which is sorted, the selected members of

a group may have something in common. For example, a researcher conducting a study on teachers training quality will choose teachers who have attended the training sessions and have sufficient knowledge of the subject matter.

11.5.4 Written Documents

This is also called record keeping method where researcher makes use of the already existing reliable documents and similar sources of information as the data source. It can include newspapers, magazines, books, websites, memos, transcripts of conversations, annual reports, and so on. Usually written documents are analysed with some form of content analysis.The advantages of using this method is that the researcher will gain better understanding of the field or subject being looked into and at the same time can get different perspectives or points of views on similar subject allowing comparisons and contrasts to be made. The drawback is that it relies heavily on the quality of the document that will be used, and the ability of the data collector to choose the right and reliable documents. The selection of wrong/irrelevant documents may affect the quality of the data collected.

11.5.5 Other Methods

The other methods that may be considered by researcher for collecting qualitative data includes Rapid assessment procedure (RAP), Free listing, Pile sort, life history (biography). These methods are discussed here in detail:

- **Rapid assessment procedure (RAP):** It is a data collection tool that provides systematic methodology for conducting rapid qualitative assessments of local conditions, needs, knowledge, attitudes and practices while taking into account the standard anthropological techniques and adapts them for use with health programmes.
- **Free listing:** It is a technique to gather data about a specific topic by asking people to list all the items they can think or relate to that topic. It is generally used to study cultural domains. For e.g. listing 5 most useful features people look for in their cars can depict the overall preference of individuals regarding car purchase.
- **Pile sort:** It is also called card sorting. Under this method, individuals are asked to sort the cards, each containing the name of an items into piles so that item in each pile are more or less similar than items in other piles. It can be very useful to explore and contextualise the relationships between individual and group norms, values, feelings etc.
- **Life history (biography):**Life-history is a method of qualitative research, frequently, but not exclusively, used in anthropology and in the health sciences today. It provides an alternative to empirical methods for identifying and documenting health patterns of individuals and groups. Under this method, people are asked to document their life over a period of time. It is a personal account of their life, in their own words and using their own personal time lines.

Thus there are multiple methods that may be appropriate for collecting data for the approach we have been chosen. The researcher in this context will need to spend some time considering which method is the best fit. Factors that should be considered when selecting a method include the objectives of the research, feasibility, time, funding, logistics, experience of research team, the type of participants and so forth.

11.6 QUALITATIVE DATA ANALYSIS: APPROACHES AND TECHNIQUES

11.6.1 Data Analysis Approaches

Qualitative Data Analysis (QDA) is the range of processes and procedures whereby we move from the qualitative data that have been collected into some form of explanation, understanding or interpretation of the people and situations we are investigating. Qualitative Data Analysis is usually based on an interpretative philosophy. The idea is to examine the meaningful and symbolic content of qualitative data. These approaches describe, either explicitly or implicitly, the purpose of the qualitative research, the role of the researcher(s), the stages of research, and the method of data analysis. Although, there are no set ground rules for analysing qualitative data, it all begins with understanding the two main approaches to qualitative data.

Deductive Approach	Inductive Approach
Used when qualitative research is a smaller component of a larger quantitative study.	Used when qualitative research is a major design of the inquiry.
The deductive approach involves analysing qualitative data based on a structure that is predetermined by the researcher. In this case, a researcher can use the questions as a guide for analysing the data.	The inductive approach, on the contrary, is not based on a predetermined structure or set ground rules/framework.
Using your research questions to group the data and then look for similarities and differences.	Using emergent framework to group the data and then look for relationships.
This approach is quick and easy and can be used when a researcher has a fair idea about the likely responses that he/she is going to receive from the sample population.	Inductive approach is often used when a researcher has very little or no idea of the research phenomenon.

Used when time and resources are limited.	This is more time consuming and a thorough approach to qualitative data analysis.

11.6.2 Data Analysis Techniques

As already discussed in above sections, qualitative data refers to non-numeric information such as interview transcripts, notes, video and audio recordings, images and text documents. Qualitative data analysisis simply the process of examining qualitative data to derive an explanation for a specific phenomenon. Qualitative data analysis gives you an understanding of your research objective by revealing patterns and themes in your data. Qualitative data analysis can be divided into the following categories:

11.6.2.1 Content Analysis

Content analysis refers to the process of categorising verbal or behavioural data for the purpose of classification, summarisation and tabulation of the data. The content can be analysed on two levels Descriptive: What is the data? Interpretative: what was meant by the data?

11.6.2.2 Narrative Analysis

Narrative analysis refers to transcribed experiences. The core activity in narrative analysis is to reformulate stories presented by people in different contexts and based on their different experiences. Every interview/ observation has narrative aspect — the researcher has to sort-out and reflect upon them, enhance them, and present them in a revised shape to the reader. Therefore, the idea of this technique is to describe the lives of individual(s) to get meaning from them. In other words, narrative analysis is the revision of primary qualitative data by researcher.

11.6.2.3 Discourse Analysis

A method of analysing a naturally occurring talk (spoken interaction) and all types of written texts. This technique focuses on ordinary people's method of producing and making sense of everyday social life. Sometimes people express themselves in a simple and straightforward way, while sometimes people express themselves vaguely and indirectly. Therefore, researcher must refer to the context when interpreting the message under this technique because the same phenomenon can be described in a number of different ways depending on context.

11.6.2.4 Framework Analysis

This is more advanced method that consists of several stages such as familiarisation, identifying a thematic framework, coding, charting, mapping and interpretation. This analysis technique involves following steps:

- Familiarisation: Transcribing and reading the data.

- Identifying a thematic framework: Initial coding framework which is developed both from a priori issues and from emergent issues.
- Coding: Using numerical or textual codes to identify specific piece of data which correspond to different themes.
- Charting: Charts created using headings from thematic framework (can be thematic or by case).
- Mapping and interpretation: Searching for patterns, associations, concepts and explanations in the data.

11.6.2.5 Grounded Theory Analysis

This method of qualitative data analysis starts with an analysis of a single case from a 'pre-defined' population in order to formulate a general statement about a population, a concept or a hypothesis. Then, additional cases are examined to see if they contribute to the theory. This method is only for limited set of analytic problems: those that can be solved with some general overall statement.

11.6.2.6 Thematic Analysis

This is one of the most common forms of analysis in qualitative research. Thematic analysis deals in pinpointing, examining, and recording themes within data. The purpose of this method is to identify the patterns of meaning across a dataset that provides an answer to the research question being addressed. Themes are patterns across data sets that are important to the description of a phenomenon and are associated to a specific research question. Most researchers consider thematic analysis to be a very useful method in capturing the intricacies of meaning within a data set.

11.7 QUALITATIVE DATA ANALYSIS: PROCEDURE AND COMPUTER SOFTWARES

11.7.1 Data Analysis Procedure

Qualitative data analysis can be conducted in following steps:

1) **Transcribe Data:** Once data is collected which is largely unstructured and sometimes makes no sense, it therefore becomes essential to first transcribe the data collected. Transcribing data means converting all the data into a textual format. For transcribing the data, researchers can either export the data into a spreadsheet or manually type the data or choose from any of the computers assisted qualitative data analysis tools e.g. CAQDAS, ATLAS.ti, NVivo, & EvaSys.

2) **Organising Data:** After transforming and arranging the data in textual format, the immediate next step is to organise the data. One great way to organise the data is going back to the research objectives/ questions and then organising the collected data based on these objectives/ questions. For e.g. arranging the research objectives into the table and assigning data according to each objective so it appears visually clear.

3) **Coding of Data:** Coding is one of the best ways to compress a huge amount of collected information. It simply means categorising and assigning concepts, properties and patterns to the collected data. A 'code' can be a word or a short phrase that represents a theme or an idea. All codes need to be assigned meaningful titles. Under this step, a wide range of non-quantifiable elements such as events, behaviours, activities, meanings etc. can be coded. Coding is an important step in qualitative data analysis as you can derive theories from relevant research findings. After assigning codes to your data, you can then begin to build on the patterns to gain in-depth insight into the data that will help make informed decisions.

Codes may be derived from the data collected, from theories, from relevant research findings or from research objectives. Some popular coding terms include:

- *Descriptive coding:* Summarising the central theme of your data.
- *In-Vivo Coding:* Using the language of the respondents to code.
- *Pattern Coding:* Finding patterns in your data and using them as the basis of your coding.

Coding can be done manually or using qualitative data analysis software such as NVivo, Atlas Ti 6.0, Hyper Research 2.8, Max QDA and others.

4) **Validation:** Validating data is one of the important steps of qualitative data analysis for successful research. Validation is all about checking the accuracy of design/methods and the second is reliability which is the extent to which your procedures produced consistent and dependable results.

5) **Conclusion of Data Analysis:** The last step is stating the findings and research outcomes based on the research objectives. There is a set of techniques that you can use to identify common themes, patterns and relationships within responses of sample group members in relation to codes that have been specified in the previous stage. Specifically, the most popular and effective methods of qualitative data interpretation used in report writing include the following:

   - Word and phrase repetitions – scanning primary data for words and phrases most commonly used by respondents, as well as, words and phrases used with unusual emotions;
   - Primary and secondary data comparisons – comparing the findings of interview/focus-group/observation/any other qualitative data collection method with the findings of literature review and discussing differences between them;
   - Search for missing information – discussions about which aspects of the issue was not mentioned by respondents, although you expected them to be mentioned;

- Metaphors and analogues – comparing primary research findings to phenomena from a different area and discussing similarities and differences.

11.7.2 Types of Software Used for Qualitative Analysis

Computer software in qualitative studies are used to transcribe data, storing data, coding data (keywords or tags), search and retrieval of data, data linking of related text, writing/editing memos about the data, display of selected reduced data, graphic mapping and preparing reports. There are various softwares available for qualitative data analysis:

- Transcribe the data (hyper TRANSCRBE software)
- Translate the data (SYSTRAN)
- Atlas Ti 6.0
- Hyper Research 2.8
- Max QDA
- The Ethnograph 5.08
- QSR N6
- QSR NVivo
- Weft QDA
- Open code 3.4

Check Your Progress 3

1) What is the difference between Interviews and Focus group Discussions?

....

....

....

....

2) Explain various approaches of Qualitative data analysis?

....

....

....

....

3) Discuss Qualitative data analysis techniques?

....

....

....

4) What are the steps involved in analysing Qualitative data?

……

……

……

11.8 REFERENCES

1) Bryman, A. (2012). *Social research methods*. Oxford university press.
2) Creswell, J. W. (2012). *Qualitative inquiry and Research Design*: Choosing among five approaches. Sage.
3) Creswell, J. W. (2013). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications, Incorporated.
4) Kothari, C. R. (2004). Research Methodology: Methods and Techniques. New Age International.
5) Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods.* SAGE Publications, inc.

11.9 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXCERCISES

Check Your Progress 1

1) See Sub-section 11.2.1 and 11.2.4

2) See Sub-section 11.3.1 and 11.3.2

3) For first part see Sub-section 11.3.3

   As in grounded theory, there is no pre-set limiting of what will be observed and no real ending point in an ethnographic study.

4) Participatory Action Research (PAR) is a qualitative research methodology due to the fact that qualitative features of an individual's feelings, views, and patterns are revealed without control or manipulation from the researcher.

5) The advantages of using case studies are:

   - It is flexible and versatile, analysing both simple and complex units and occurrence, even over a long period of time.
   - Case studies provide in-depth and detailed information, thanks to how it captures as many variables as it can.

   The disadvantage of case study is that reliability of the data may be put at risk when the case study or studies chosen are not representative of the sample or population.

Check Your Progress 2

1) See Sub-section 11.4.1

2) See Sub-section 11.4.2 and 11.4.3

3) Although there is no fixed criterion of deciding the sample size as it depends upon the research question under study. Despite this, researcher may consider two most important factors while taking sample size decisions in qualitative research. These factors are:

- The one that adequately answers the research question- until new categories, themes or explanations stop emerging from the data
- Depend on available time and resources
- The sample size may be predetermined or based on theoretical saturation, which is the point at which the newly collected no longer provides additional insights.
- Once a sampling method has been determined, the researcher must consider the sample size. In qualitative studies, sampling typically continues until information redundancy or saturation occurs. This is the point at which no new information is emerging in the data. Therefore, in qualitative studies is it critical that data collection and analysis are occurring simultaneously so that the researcher will know when the saturation point is reached. It is important to understand that the saturation point may occur prematurely if the researcher has a narrow sampling frame, a skewed analysis of the data, or poor methodology. Because of this, the researcher must carefully create the research question, select an appropriate target group, eliminate his or her own biases and analyse data continuously and thoroughly throughout the process to bring validity to the data collected.

Check Your Progress 3

1) See Sub-sections 11.5.2 and 11.5.3

2) See Sub-section 11.6.1

3) See Sub-section 11.6.2

4) See Sub-section 11.7.1

KEY WORDS

Abscissa (X-axis) : The horizontal axis of a graph.

Array : A rough grouping of data.

A Standard Score or z Score : Tells us how many standard deviations a data value is above or below the mean for a specific distribution of values.

Alternative Hypothesis : Symbolised by H_1, is a statistical hypothesis that states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters.

Analysis of Variance (ANOVA) : A statistical technique for determining the statistical significance of differences among means; it can be used with two or more groups and uses the F-test statistic.

Bar diagram : A diagram consisting of thick vertical lines corresponding to values of variables.

Body of the Table : This is the real table and contains numerical information or data in different cells.

Bootstrapping : An approach to validating a multivariate model by drawing a large number of sub-samples and estimating the model for each sample. Estimates from all the subsamples, providing not only the best estimated coefficients but their expected variability and thus their likelihoods of differing from zero; that is are the estimated coefficients are statistically different from zero or not?

This approach does not rely on the statistical assumption about the population to assess statistical significance but instead makes its assessment based solely on the sample data.

Bias : Systematic difference between predicted and measured values. The bias is computed as the average value of the residuals.

Central tendency : Tendency of the distribution to converge towards the centre.

Census : Collection of information/data from or about all the unity in the universe.

Caption : Titles given to the columns of a table are called captions.

Classification : A systematic grouping of data.

Continuous : When data are in regular in a classification.

Cumulative Frequency : Cumulative frequency of a class is obtained by adding its frequency and the sum of frequencies of all the classes preceding it.

Correlation : A measure of linear association between two (ordered) lists. Two variables can be strongly correlated without having any causal relationship, and two variables can have a causal relationship and yet be uncorrelated.

Correlation Coefficient : Measures the strength and direction of a linear relationship between two variables.

Coefficient of Determination : The coefficient of determination is the ratio of the explained variation to the total variation and is equal to r^2. The predictability of a regression equation could also be judged by its coefficient of determination.

Confidence Interval : A confidence interval for a parameter is a random interval constructed from data in such a way that the probability that the interval contains the true value of the parameter can be specified before the data are collected.

Confidence Level : The confidence level of a confidence interval is the chance that the interval that will result once data are collected will contain the corresponding parameter. If one computes confidence intervals again and again from independent sample, the long-term limit of the fraction of intervals that contain the parameter is the confidence level.

Coding Sorts : Compilation of similarly coded blocks of text from different sources into a single file or report.

Coding : The process of attaching labels to lines of text so that the researcher can group and compare similar or related pieces of information.

Descriptive Statistics : Branch of statistics which describe the basic features of the data in a study.

Dispersion : Degree of spread out on either side of the center.

Data : The information collected from different

Key words

sources through the use of various tools and techniques comprising of numerical figures, ratings, descriptive narratives, field notes etc., are called data. In short, data refers to any sort of information that can be analysed.

Discrete : When data are counted in a classification.

Degrees of Freedom : The number of values that are free to vary after a sample statistic has been computed.

Dependent Variable : In regression, the variable whose values are supposed to be explained by changes in the other variable (the independent or explanatory variable). Usually one regresses the dependent variable on the independent variable.

Data Mining : This is the practice of studying large amounts of data to find patterns or trends. MVA is a form of data mining.

Epistemology : Epistemology refers to the theory of knowledge of how human beings come to have knowledge of the world around them of how we know what we know. Broadly there are two theories: Rationalism and Empiricism.

Exclusive Classification : The classification system in which the upper limit of the class becomes the Lower limit of next class.

Estimator The Sample measures (i.e., statistics) used to estimate population measures (i.e. parameters).

Explanatory Variable : In regression, the explanatory or independent variable is the one that is supposed to "explain" the other. For example, in examining crop yield versus quantity of fertilizer applied, the quantity of fertilizer would be the explanatory or independent variable, and the crop yield would be the dependent variable. In experiments, the explanatory variable is the one that is manipulated; the one that is observed is the dependent variable.

Evaluation Research : The researcher provides a causal explanation of 'why it is so?' or a functional explanation of 'how is it so?' for a phenomenon under study.

Experimental Design : This is also referred to as Design of Experiments. Plan for experiments where input variables are varied systematically within predefined ranges, so that their effects on the output variables (responses) can be estimated and checked for significance. Experimental designs are built with a specific objective in mind, namely screening, screening with interaction, or optimisation. The number of experiments and the way they are built depends on the objective and on the operational constraints.

Focus Group Discussion : It is a technique for a eliciting descriptive information/ data from specific population subgroups (a group of 8 to 12 persons).

Frequency Distribution : Arrangement of the frequency of individual values or ranges of values for a variable.

Frequency Polygon : It is a broken line graph to represent frequency distribution.

Five-Number Summary : Minimum value, median, Q_1, Q_3 and max value in a distribution used to draw boxplot to represent the data set to depict the symmetry of the distribution.

Histogram : It is a set of adjacent rectangles presented vertically with areas proportional to the frequencies of the corresponding classes.

Hotelling T^2 Statistics : A linear function of the leverage that can be compared to a critical limit according to an F-test. This statistic is useful for the detection of outliers at the modeling or prediction stage.

Iteration : It is an act of repeating a process with the aim of approaching a desired goal, target or result.

Inclusive Classification : When the upper limit of a class differs from the lower limit of its successive class.

Inferential Statistics : Inferential statistics deals with drawing inferences about a population from samples drawn from it. Two important features of inferential statistics are estimating parameters and hypothesis testing.

Independent Variable : In regression, the independent variable is the one that is supposed to explain the

other; the term is a synonym for "explanatory variable". Usually, one regresses the "dependent variable" on the "independent variable". There is not always a clear choice of the “independent variable”. The independent variable is usually plotted on the horizontal axis. Independent in this context does not mean the same thing as statistically independent.

Indexing : Process that generates a word list comprising all the substantive words and their location within the texts entered into a program.

Multivariate Analysis : Analysis of multiple variables in a single relationship or a set of relationships.

Measurement Error : Inaccuracies of measuring the true variable value due to the fallibility of measurement instrument (inappropriate scales), due to entry errors or respondent errors.

Multicollinearity : Multicollinearity represents the degree to which any variable’s effect can be predicted or accounted for by the other variables in analysis. As multicollinearity rises, the ability to define any variable effect is diminished. The addition of irrelevant variable can only increase the degree of multicollinearity which makes the interpretation all the more difficult.

Model : Mathematical equation summarising variations in a data set. Models are built so that the structure of a data table can be understood better than by just looking at all raw values. Statistical models consist of a structure part and an error part. The structure part (information) is intended to be used for interpretation or prediction, and the error part (noise) should be as small as possible for the model to be reliable.

Non-Sampling Error : Errors occurring due to faulty measurement, false data, biased attitude of investigator, memory lapse on the part of respondent, faulty sampling frame, non-response or faulty processing and presentation.

Null Hypothesis : Symbolised by $H_{o,}$ is a statistical hypothesis that states that there is no difference between a parameter and a specific value,

or that there is no difference between two parameters.

Noise : Random variation that does not contain any information. The purpose of multivariate modelling is to separate information from noise.

Ogive : It is the graph of cumulative frequency.

Open-end distributions : Classification on which one or more classes are open-ended.

Ordinate (Y-axis) : The vertical axis of a graph.

Outlier : An observation that deviates from the mean many times of standard deviation. It sometimes tempts to discard outliers, but this is imprudent unless the cause of the outlier can be identified, and the outlier is determined to be spurious. Otherwise, discarding outliers can cause one to underestimate the true variability of the measurement process.

Overfitting : For a model, overfitting is a tendency to describe too much of the variation in the data, so that not only consistent structure is taken into account, but also some noise or non-informative variation. Overfitting should be avoided, since it usually results in a lower quality of prediction. Validation is an efficient way to avoid model overfitting.

Orthogonal : Two variables are said to be orthogonal if they are completely uncorrelated, i.e. their correlation is 0. In PCA and PCR, the principal components are orthogonal to each other.

Population or Universe : Population or universe refers to the complete set or collection of animate beings or inanimate objects which are subject of an enquiry or study.

Primary Data : The primary data are kind of data that are collected freshly for the first time, and thus it is original in nature.

Pictogram : In pictogram data are presented in the form of pictures.

Pie Diagram : It is a circle sub-divided into components to present proportion of different constituent parts of a total.

Key words

Primary Data	:	The information gathered direct from the person or entity/unit about which the information is required.
Parameter	:	An unknown numerical summary of the population.
Power	:	The power of a test against a specific alternative hypothesis is the chance that the test correctly rejects the null hypothesis when the alternative hypothesis is true.
Principal Component Analysis (PCA)	:	PCA is a bilinear modeling method which gives an interpretable overview of the main information in a multidimensional data table. The information carried by the original variables is projected onto a smaller number of underlying ("latent") variables called principal components. The first principal component covers as much of the variation in the data as possible. The second principal component is orthogonal to the first and covers as much of the remaining variation as possible, and so on.
Practical Significance	:	Means of assessing multivariate analysis results based on their substantive findings rather than their statistical significance. Whereas statistical significance determines whether the result is attributable to chance, practical significance assesses whether the result is useful in achieving the research objectives.
Questionnaire	:	The questionnaire is a research instrument consisting of a series of questions related to the research topic in written or printed form, and respondents have to answer these questions in the space meant for it.
Qualitative Classification	:	When data are classified on the basis of attributes.
Quantitative Classification	:	When data are classified on the basis of number or frequency.
Reflexivity	:	Reflexivity refers to a situation where an individual becomes a self by being able to take the attitude of others and thereby reflect on his or her own behaviour.
Relative Frequency Distribution	:	It is a frequency distribution where the frequency of each value is expressed as a fraction or percentage of the total number

of observations.

R-square : The R-square of a regression model is a measure of the quality of the model. Also known as coefficient of determination, it is computed as 1 – (Residual Y-variance)/ Total variance, or (Explained Y-variance)/Total variance. For Calibration results, this is also the square of the correlation coefficient between predicted and measured values, and the R-square value is always between 0 and 1. The closer to 1, the better.

Reliability : Extent to which a variable or set of variables is consistent in what is intended to measure. It differs from validity in that it relates not to what should be measures, but instead how it is measured.

Sample Size : Refers to the number of units in a sample.

Sampling Frame : The list of the units of the universe used for selecting a sample.

Secondary Data : The secondary data are kind of data that is already collected by other researcher and the data have already been gone through the process of statistical analysis.

Secondary Data : Information gathered through already maintained records about a variable.

Structured Interview : Structured interviews comprise of a fixed set of predetermined questions and standardised methods record the responses.

Schedules : The schedule is a sort of proforma, which contains a list of questions to be filled by the researchers.

Stub : It is a part of table. It stands for brief and self explanatory headings of rows.

Statistic : A known numerical summary of the sample which can be used to make inference about parameters.

Significance : The measure of whether the results of research were due to chance. The more the statistical significance assigned to an observation, the less likely the observation occurred by chance.

Summed Scales : Method of combining several variables that measure the same concept into a single variable in an attempt to increase the

Key words

reliability of measurement through multivariate measurement. In most instances, the separate variables are summed up and their total or average score is used in the analysis.

Screening : First stage of an investigation, where information is sought about the effects of many variables. Since many variables have to be investigated, only main effects, and optionally interactions, can be studied at this stage.

Tabulation : It is a systematic presentation of classified data in rows and columns with Appropriate headings and sub headings.

t-test : A hypothesis test based on approximating the probability histogram of the test statistic by Student's t curve. t-tests usually are used to test hypotheses about the mean of a population or difference of means of two population when the sample size is intermediate and the distribution of the population is known to be nearly normal.

Transcription : Many qualitative studies collect audio or video data (e.g. recordings of interviews, focus groups or talk in consultation), and these are usually transcribed into written form for closer study.

Theory : A set of interrelated concepts, definitions and propositions that presents a systematic view of events or situations by specifying relations among variables.

Themes : Idea categories that emerge from grouping of lower-level data points.

Unstructured Interviews : Unstructured interviews comprise of the flexible approach in terms of asking questions. It does not follow a system of predetermined questions and standardised techniques of recording responses.

Validity : Extent to which a measure or set of measures correctly represents the concept of study – the degree to which it is free from any systematic error. Validity is concerned with how well the concept is defined by the measures(s) whereas reliability relates to the consistency of the measures(s).

Variate : A linear combination of variables with empirically determined weights.

Weighting : A technique to modify the relative influences of the variables on a model. This is achieved by giving each variable a new weight, i.e. multiplying the original values by a constant which differs between variables. This is also called scaling.

z-test : A hypothesis test based on approximating the probability histogram of the Z statistic under the null hypothesis by the normal curve.

ignou
THE PEOPLE'S
UNIVERSITY

SOME USEFUL BOOKS

1. Asthana, H.S. and Bhushan, B. (2016): *Statistics for Social Sciences (with SPSS Application).* Second Edition, Prentice Hall of India, New Delhi.
2. Donald H. Sanders (1994): *Statistics A First Course:* 6th Edition, Macro Hill Publications.
3. Damodar Gujrati (2016): *Econometrics by examples,* Palgrave Publications.
4. M.R. Spiegel (2003): *Theory and problems of probability and statistics* (Schaum Series).
5. P.H. Karmel and M. Polasek (1978): *Applied statistics for Economists,* 4th Edition, Pitman

ignou THE PEOPLE'S UNIVERSITY

MPDD/IGNOU/P.O. 7 K/ January, 2021

ISBN: 978-93-90773-19-0