Module III: Data Analysis – II:

Factor analysis, Multiple Regressions Analysis. Discriminant Analysis (Concept)

Report writing and presentation: Research Report, Types and significance, Structure of research report, Presentation of report.

FACTOR ANALYSIS:

There are several methods of factor analysis, but they do not necessarily give same results. As such factor analysis is not a single unique method but a set of techniques. Important methods of factor analysis are:

(i) The centroid method;

(ii) The principal components method;

(ii) The maximum likelihood method.

Before we describe these different methods of factor analysis, it seems appropriate that some basic terms relating to factor analysis be well understood.

(i) ***Factor:*** A factor is an underlying dimension that account for several observed variables. There can be one or more factors, depending upon the nature of the study and the number of variables involved in it.

(ii) ***Factor-loadings:*** Factor-loadings are those values which explain how closely the variables are related to each one of the factors discovered. They are also known as factor-variable correlations.

In fact, factor-loadings work as key to understanding what the factors mean. It is the absolute size (rather than the signs, plus or minus) of the loading that is important in the interpretation of a factor.

(iii) ***Communality*** **(h^2)*:*** Communality, symbolized as *h*2, shows how much of each variable is accounted for by the underlying factor taken together. A high value of communality means that not much of the variable is left over after whatever the factors represent is taken into consideration. It is worked out in respect of each variable as under:

$$h^2 \text{ of the } i\text{th variable} = (i\text{th factor loading of factor } A)^2 + (i\text{th factor loading of factor } B)^2 + \ldots$$

(iv) *Eigen value (or latent root):* When we take the sum of squared values of factor loadings relating to a factor, then such sum is referred to as Eigen Value or latent root. Eigen value indicates the relative importance of each factor in accounting for the particular set of variables being analysed.

(v) *Total sum of squares:* When eigen values of all factors are totalled, the resulting value is termed as the total sum of squares. This value, when divided by the number of variables (involved in a study), results in an index that shows how the particular solution accounts for what all the variables taken together represent. If the variables are all very different from each other, this index will be low. If they fall into one or more highly redundant groups, and if the extracted factors account for all the groups, the index will then approach unity.

(vi) *Rotation:* Rotation, in the context of factor analysis, is something like staining a microscope slide. Just as different stains on it reveal different structures in the tissue, different rotations reveal different structures in the data. Though different rotations give results that appear to be entirely different, but from a statistical point of view, all results are taken as equal, none superior or inferior to others. However, from the standpoint of making sense of the results of factor analysis, one must select the right rotation. If the factors are independent orthogonal rotation is done and if the factors are correlated, an oblique rotation is made. Communality for each variable will remain undisturbed regardless of rotation but the Eigen values will change as result of rotation.

(vii) *Factor scores:* Factor score represents the degree to which each respondent gets high scores on the group of items that load high on each factor. Factor scores can help explain what the factors mean. With such scores, several other multivariate analyses can be performed.

We can now take up the important methods of factor analysis.

(A) Centroid Method of Factor Analysis:

This method of factor analysis, developed by L.L. Thurstone, was quite frequently used until about 1950 before the advent of large capacity high speed computers. The centroid method tends to maximize the sum of loadings, disregarding signs; it is the method which extracts the largest sum of absolute loadings for each factor in turn. It is defined by linear combinations in which all weights are either + 1.0 or – 1.0. The main merit of this method is that it is relatively simple, can be easily understood and involves simpler computations. If one understands this method, it becomes easy to understand the mechanics involved in other methods of factor analysis.

Various steps involved in this method are as follows:

(i) This method starts with the computation of a matrix of correlations, *R*, wherein unities are place in the diagonal spaces. The product moment formula is used for working out the correlation coefficients.

(ii) If the correlation matrix so obtained happens to be positive manifold (i.e., disregarding the diagonal elements each variable has a large sum of positive correlations than of negative correlations), the centroid method requires that the weights for all variables be +1.0. In other words, the variables are not weighted; they are simply summed. But in case the correlation matrix is not a positive manifold, then reflections must be made before the first centroid factor is obtained.

(iii) The first centroid factor is determined as under:

(a) The sum of the coefficients (including the diagonal unity) in each column of the correlation matrix is worked out.

(b) Then the sum of these column sums (*T*) is obtained.

(c) The sum of each column obtained as per (a) above is divided by the square root of *T* obtained in (b) above, resulting in what are called centroid loadings. This way each centroid loading (one loading for one variable) is computed. The full set of loadings so obtained constitute the first centroid factor (say *A*).

(iv) To obtain second centroid factor (say *B*), one must first obtain a matrix of residual coefficients. For this purpose, the loadings for the two variables on the first centroid factor are multiplied. This is done for all possible pairs of variables (in each diagonal space is the square of the particular factor loading). The resulting matrix of factor cross products may be named as Q_1. Then Q_1 is subtracted clement by element from the original matrix of correlation, *R*, and the result is the first matrix of residual coefficients, R_1. After obtaining *R*1, one must *reflect* some of the variables in it, meaning thereby that some of the variables are given negative signs in the sum [This is usually done by inspection. The aim in doing this should be to obtain a reflected matrix, R'_1, which will have the highest possible sum of coefficients (*T*)]. For any variable which is so reflected, the signs of all coefficients in that column and row of the residual matrix are changed. When this is done, the matrix is named as 'reflected matrix' form which the loadings are obtained in the usual way (already explained in the context of first centroid factor), but the loadings of the variables which were reflected must be given negative signs. The full set of loadings so obtained constitutes the second centroid factor (say *B*). Thus loadings on the second centroid factor are obtained from R'_1.

(v) For subsequent factors (*C*, *D*, etc.) the same process outlined above is repeated. After the second centroid factor is obtained, cross products are computed forming, matrix, Q_2. This is then subtracted from R_1 (and not from R'_1) resulting in R_2. To obtain a third factor (*C*), one should operate on R_2 in the same way as on R_1. First, some of the variables would have to be reflected to maximize the sum of loadings, which would produce R'_2. Loadings would be computed from R'_2 as they were from R'_1. Again, it would be necessary to give negative signs to the loadings of variables which were reflected which would result in third centroid factor (*C*).

EXAMPLE:

Given is the following correlation matrix, *R*, relating to eight variables with unities in the diagonal spaces:

Variables	1	2	3	4	5	6	7	8
1	1.000	.709	.204	.081	.626	.113	.155	.774
2	.709	1.000	.051	.089	.581	.098	.083	.652
3	.204	.051	1.000	.671	.123	.689	.582	.072
4	.081	.089	.671	1.000	.022	.798	.613	.111
5	.626	.581	.123	.022	1.000	.047	.201	.724
6	.113	.098	.689	.798	.047	1.000	.801	.120
7	.155	.083	.582	.613	.201	.801	1.000	.152
8	.774	.652	.072	.111	.724	.120	.152	1.000

Using the centroid method of factor analysis, work out the first and second centroid factors from the above information.

Solution: Given correlation matrix, R, is a positive manifold and as such the weights for all variables be +1.0. Accordingly, we calculate the first centroid factor (A) as under:

Table 13.1(a)

Variables

		1	2	3	4	5	6	7	8
	1	1.000	.709	.204	.081	.626	.113	.155	.774
	2	.709	1.000	.051	.089	.581	.098	.083	.652
	3	.204	.051	1.000	.671	.123	.689	.582	.072
Variables	4	.081	.089	.671	1.000	.022	.798	.613	.111
	5	.626	.581	.123	.022	1.000	.047	.201	.724
	6	.113	.098	.689	.798	.047	1.000	.801	.120
	7	.155	.083	.582	.613	.201	.801	1.000	.152
	8	.774	.652	.072	.111	.724	.120	.152	1.000
Column sums		3.662	3.263	3.392	3.385	3.324	3.666	3.587	3.605

Sum of the column sums (T) = 27.884 $\quad \therefore \sqrt{T} = 5.281$

First centroid factor $A = \frac{3.662}{5.281}, \frac{3.263}{5.281}, \frac{3.392}{5.281}, \frac{3.385}{5.281}, \frac{3.324}{5.281}, \frac{3.666}{5.281}, \frac{3.587}{5.281}, \frac{3.605}{5.281}$

= .693, .618, .642, .641, .629, .694, .679, .683

We can also state this information as under:

Table 13.1 (b)

Variables	*Factor loadings concerning first Centroid factor A*
1	.693
2	.618
3	.642
4	.641
5	.629
6	.694
7	.679
8	.683

To obtain the second centroid factor B, we first of all develop (as shown on the next page) the first matrix of factor cross product, Q_1:

First Matrix of Factor Cross Product (Q_1)

First centroid factor *A*

	.693	.618	.642	.641	.629	.694	.679	.683
.693	.480	.428	.445	.444	.436	.481	.471	.473
.618	.428	.382	.397	.396	.389	.429	.420	.422
.642	.445	.397	.412	.412	.404	.446	.436	.438
.641	.444	.396	.412	.411	.403	.445	.435	.438
.629	.436	.389	.404	.403	.396	.437	.427	.430
.694	.481	.429	.446	.445	.437	.482	.471	.474
.679	.471	.420	.436	.435	.427	.471	.461	.464
.683	.473	.422	.438	.438	.430	.474	.464	.466

Now we obtain first matrix of residual coefficient (R_1) by subtracting Q_1 from *R* as shown below:

First Matrix of Residual Coefficient (R_1)

Variables

Variables	1	2	3	4	5	6	7	8
1	.520	.281	−.241	−.363	.190	−.368	−.316	.301
2	.281	.618	−.346	−.307	.192	−.331	−.337	.230
3	−.241	−.346	.588	.259	−.281	.243	.146	−.366
4	−.363	−.307	.259	.589	−.381	.353	.178	−.327
5	.190	.192	−.281	−.381	.604	−.390	−.217	.294
6	−.368	−.331	.243	.353	−.390	.518	.330	−.354
7	−.316	−.337	.146	.178	−.226	.330	.539	−.312
8	.301	.230	−.366	−.327	.294	−.354	−.312	.534

Reflecting the variables 3, 4, 6 and 7, we obtain reflected matrix of residual coefficient (R'_1) as under and then we can extract the second centroid factor (*B*) from it as shown on the next page.

Reflected Matrix of Residual Coefficients (R'_1)
and Extraction of 2nd Centroid Factor (*B*)

Variables

	1	2	3*	4*	5	6*	7*	8
Variables 1	.520	.281	.241	.363	.190	.368	.316	.301
2	.281	.618	.346	.307	.192	.331	.337	.230
3*	.241	.346	.588	.259	.281	.243	.146	.366
4*	.363	.307	.259	.589	.381	.353	.178	.327
5	.190	.192	.281	.381	.604	.390	.217	.294
6*	.368	.331	.243	.353	.390	.518	.330	.354
7*	.316	.337	.146	.178	.226	.330	.539	.312

Contd.

Variables

	1	2	3*	4*	5	6*	7*	8
8	.301	.230	.366	.327	.294	.354	.312	.534
Column sums:	2.580	2.642	2.470	2.757	2.558	2.887	2.375	2.718
Sum of column sums (T) = 20.987 ∴ $\sqrt{T} = 4.581$								
Second centroid factor B = .563 .577 –.539 –.602 .558 –.630 –.518 .593								

*These variables were reflected.

Now we can write the matrix of factor loadings as under:

Variables	*Factor loadings*	
	Centroid Factor A	*Centroid Factor B*
1	.693	.563
2	.618	.577
3	.642	–.539
4	.641	–.602
5	.629	.558
6	.694	–.630
7	.679	–.518
8	.683	.593

(B) Principal-components Method of Factor Analysis:

Principal-components method (or simply P.C. method) of factor analysis, developed by H. Hotelling, seeks to maximize the sum of squared loadings of each factor extracted in turn. Accordingly PC factor explains more variance than would the loadings obtained from any other method of factoring. The aim of the principal components method is the construction out of a given set of variables X_j's ($j = 1, 2, \ldots, k$) of new variables (p_i), called principal components which are linear combinations of the X_s

$$p_1 = a_{11} X_1 + a_{12} X_2 + \ldots + a_{1k} X_k$$
$$p_2 = a_{21} X_1 + a_{22} X_2 + \ldots + a_{2k} X_k$$
$$\vdots$$
$$p_k = a_{k1} X_1 + a_{k2} X_2 + \ldots + a_{kk} X_k$$

The method is being applied mostly by using standardized variables, i.e., $z_j = \left(X_j - \overline{X}_j\right)^2 / \sigma_j$.

The *aij*'s are called loadings and are worked out in such a way that the extracted principal components satisfy two conditions: (i) principal components are uncorrelated (orthogonal) and (ii) the first principal component (p_1) has the maximum variance, the second principal component (p_2) has the next maximum variance and so on.

Following steps are usually involved in principal components method

(i) Estimates of a_{ij}'s are obtained with which X's are transformed into orthogonal variables i.e., the principal components. A decision is also taken with regard to the question: how many of the components to retain into the analysis?

(ii) We then proceed with the regression of Y on these principal components i.e.,

$$Y = \hat{y}_1 p_1 + \hat{y}_2 p_2 + \ldots + \hat{y}_m p_m \ (m < k)$$

(iii) From the $\hat{a}_{ij}$ and $\dot{y}_{ij}$, we may find b_{ij} of the original model, transferring back from the p's into the standardized X's.

Alternative method for finding the factor loadings is as under:

(i) Correlation coefficients (by the product moment method) between the pairs of k variables are worked out and may be arranged in the form of a correlation matrix, R, as under:

Correlation Matrix, R

Variables

		X_1	X_2	X_3	....	X_k
	X_1	r_{11}	r_{12}	r_{13}	....	r_{1k}
	X_2	r_{21}	r_{22}	r_{23}	....	r_{3k}
Variables	X_3	r_{31}	r_{32}	r_{33}	....	r_{3k}
	.	.	.	.		
	.	.	.	.		
	X_k	r_{k1}	r_{k2}	r_{k3}	....	r_{kk}

The main diagonal spaces include unities since such elements are self-correlations. The correlation matrix happens to be a symmetrical matrix.

(ii) Presuming the correlation matrix to be positive manifold (if this is not so, then reflections as mentioned in case of centroid method must be made), the first step is to obtain the sum of coefficients in each column, including the diagonal element. The vector of column sums is referred to as U_{a1} and when U_{a1} is normalized, we call it V_{a1}. This is done by squaring and summing the column sums in U_{a1} and then dividing each element in U_{a1} by the square root of the sum of squares (which may be termed as normalizing factor). Then elements in V_{a1} are accumulatively multiplied by the first row of R to obtain the first element in a new vector U_{a2}. For instance, in multiplying V_{a1} by the first row of R, the first element in V_{a1} would be multiplied by the r_{11} value and this would be added to the product of the second element in V_{a1} multiplied by the r_{12} value, which would be added to the product of third element in V_{a1} multiplied by the r_{13} value, and so on for all the corresponding elements in V_{a1} and the first row of R. To obtain the second element of U_{a2}, the same process would be repeated i.e., the elements in V_{a1} are accumulatively multiplied by the 2nd row of R. The same process would be repeated for each row of R and the result would be a new vector U_{a2}. Then U_{a2} would be normalized to obtain V_{a2}. One would then compare V_{a1} and V_{a2}. If they are nearly identical, then convergence is said to have occurred (If convergence does not occur, one should go on using these trial vectors again and again till convergence occurs). Suppose the convergence occurs when we work out V_{a8} in which case V_{a7} will be taken as V_a (the characteristic vector) which can be converted into loadings on the first principal component when we multiply the said vector (i.e., each element of V_a) by the square root of the number we obtain for normalizing U_{a8}.

(iii) To obtain factor B, one seeks solutions for V_b, and the actual factor loadings for second component factor, B. The same procedures are used as we had adopted for finding the first factor, except that one operates off the first residual matrix, R_1 rather than the original correlation matrix R (We operate on R_1 in just the same way as we did in case of centroid method stated earlier).

(iv) This very procedure is repeated over and over again to obtain the successive PC factors (viz. C, D, etc.).

Other steps involved in factor analysis

(a) Next the question is: How many principal components to retain in a particular study? Various criteria for this purpose have been suggested, but one often used is Kaiser’s criterion. According to this criterion only the principal components, having latent root greater than one, are considered as essential and should be retained.

(b) The principal components so extracted and retained are then rotated from their beginning position to enhance the interpretability of the factors.

(c) Communality, symbolized, h^2, is then worked out which shows how much of each variable is accounted for by the underlying factors taken together. A high communality figure means that not much of the variable is left over after whatever the factors represent is taken into consideration. It is worked out in respect of each variable as under:

$$h^2 \text{ of the } i\text{th variable} = (i\text{th factor loading of factor } A)^2$$
$$+ (i\text{th factor loading of factor } B)^2 + \dots$$

Then follows the task of interpretation. The amount of variance explained (sum of squared loadings) by each PC factor is equal to the corresponding characteristic root. When these roots are divided by the number of variables, they show the characteristic roots as proportions of total variance explained.

(d) The variables are then regressed against each factor loading and the resulting regression coefficients are used to generate what are known as factor scores which are then used in further analysis and can also be used as inputs in several other multivariate analyses.

EXAMPLE:

Take the correlation matrix, R, for eight variables of illustration 1 of this chapter and then compute:

(i) The first two principal component factors;

(ii) The communality for each variable on the basis of said two component factors;

(iii) The proportion of total variance as well as the proportion of common variance explained by each of the two component factors.

Solution: Since the given correlation matrix is a positive manifold, we work out the first principal component factor (using trial vectors) as under:

		Variables							
		1	2	3	4	5	6	7	8
	1	1.000	.709	.204	.081	.626	.113	.155	.774
	2	.709	1.000	.051	.089	.581	.098	.083	.652
	3	.204	.051	1.000	.671	.123	.689	.582	.072
	4	.081	.089	.671	1.000	.022	.798	.613	.111
Variables	5	.626	.581	.123	.022	1.000	.047	.201	.724
	6	.113	.098	.689	.798	.047	1.000	.801	.120
	7	.155	.083	.582	.613	.201	.801	1.000	.152
	8	.774	.652	.072	.111	.724	.120	.152	1.000
Column sums U_{a1}		3.662	3.263	3.392	3.385	3.324	3.666	3.587	3.605
Normalizing U_{a1} we obtain V_{a1} i.e., V_{a1} = U_a/Normalizing factor*		.371	.331	.344	.343	.337	.372	.363	.365

*Normalizing factor $= \sqrt{(3.662)^2 + (3.263)^2 + (3.392)^2 + (3.385)^2 + (3.324)^2 + (3.666)^2 + (3.587)^2 + (3.605)^2}$

$= \sqrt{97.372} = 9.868$

Then we obtain U_{a2} by accumulatively multiplying V_{a1} row by row into R and the result comes as under:

U_{a2} : [1.296, 1.143, 1.201, 1.201, 1.165, 1.308, 1.280, 1.275]

Normalizing it we obtain (normalizing factor for U_{a2} will be worked out as above and will be = 3.493):

V_{a2} : [.371, .327, .344, .344, .334, .374, .366, .365]

Comparing V_{a1} and V_{a2}, we find the two vectors are almost equal and this shows convergence has occurred. Hence V_{a1} is taken as the characteristic vector, V_a. Finally, we compute the loadings on the first principal component by multiplying V_a by the square root of the number that we obtain for normalizing U_{a2}. The result is as under:

Variables	*(Characteristic vector V_a)*	×	$\sqrt{\textit{normalizing factor of } U_{a2}}$	=	*Principal Component I*
1	.371	×	1.868	=	.69
2	.331	×	1.868	=	.62
3	.344	×	1.868	=	.64
4	.343	×	1.868	=	.64
5	.337	×	1.868	=	.63
6	.372	×	1.868	=	.70
7	.363	×	1.868	=	.68
8	.365	×	1.868	=	.68

For finding principal component II, we have to proceed on similar lines (as stated in the context of obtaining centroid factor *B* earlier in this chapter) to obtain the following result:

Variables	*Principal Component II*
1	+.57
2	+.59
3	–.52
4	–.59
5	+.57
6	–.61
7	–.49
8	–.61

The other parts of the question can now be worked out (after first putting the above information in a matrix form) as given below:

Variables	*Principal Components*		*Communality,* h^2
	I	*II*	
1	.69	+.57	$(.69)^2+(.57)^2=.801$
2	.62	+.59	$(.62)^2+(.59)^2=.733$
3	.64	−.52	$(.64)^2+(-.52)^2=.680$
4	.64	−.59	$(.64)^2+(-.59)^2=.758$
5	.63	+.57	$(.63)^2+(.57)^2=.722$
6	.70	−.61	$(.70)^2+(-.61)^2=.862$
7	.68	−.49	$(.68)^2+(-.49)^2=.703$
8	.68	−.61	$(.68)^2+(-.61)^2=.835$
Eigen value i.e., common variance	3.4914	2.6007	6.0921
Proportion of total variance	.436 (43.6%)	.325 (32.5%)	.761 (76%)
Proportion of common variance	.573 (57%)	.427 (43%)	1.000 (100%)

All these values can be interpreted in the same manner as stated earlier.

(C) Maximum Likelihood (ML) Method of Factor Analysis:

The ML method consists in obtaining sets of factor loadings successively in such a way that each, in turn, explains as much as possible of the population correlation matrix as estimated from the sample correlation matrix. If R_s stands for the correlation matrix actually obtained from the data in a sample, R_p stands for the correlation matrix that would be obtained if the entire population were tested, then the ML method seeks to extrapolate what is known from *Rs* in the best possible way to estimate R_p (but the PC method only maximizes the variance explained in R_s). Thus, the ML method is a statistical approach in which one maximizes some relationship between the sample of data and the population from which the sample was drawn.

The arithmetic underlying the ML method is relatively difficult in comparison to that involved in the PC method and as such is understandable when one has adequate grounding in calculus, higher algebra and matrix algebra in particular. Iterative approach is employed in ML method also to find each factor, but the iterative procedures have proved much more difficult than what we find in the case of PC method. Hence the ML method is generally not used for factor analysis in practice.

The loadings obtained on the first factor are employed in the usual way to obtain a matrix of the residual coefficients. A significance test is then applied to indicate whether it would be reasonable to extract a second factor. This goes on repeatedly in search of one factor after another. One stops factoring after the significance test fails to reject the null hypothesis for the residual matrix. The final product is a matrix of factor loadings. The ML factor loadings can be interpreted in a similar fashion as we have explained in case of the centroid or the PC method.

MULTI REGRESSION ANALYSIS:

When there are two or more than two independent variables, the analysis concerning relationship is known as multiple correlations and the equation describing such relationship as the multiple regression equation. We here explain multiple correlation and regression taking only two independent variables and one dependent variable (Convenient computer programs exist for dealing with a great number of variables). In this situation the results are interpreted as shown below:

Multiple regression equation assumes the form

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

where X_1 and X_2 are two independent variables and Y being the dependent variable, and the constants a, b_1 and b_2 can be solved by solving the following three normal equations:

$$\sum Y_i = na + b_1 \sum X_{1i} + b_2 \sum X_{2i}$$
$$\sum X_{1i}Y_i = a \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i}X_{2i}$$
$$\sum X_{2i}Y_i = a \sum X_{2i} + b_1 \sum X_{1i}X_{2i} + b_2 \sum X_{2i}^2$$

(It may be noted that the number of normal equations would depend upon the number of independent variables. If there are 2 independent variables, then 3 equations, if there are 3 independent variables then 4 equations and so on, are used.)

In multiple regression analysis, the regression coefficients (viz., b_1 b_2) become less reliable as the degree of correlation between the independent variables (viz., X_1, X_2) increases. If there is a high degree of correlation between independent variables, we have a problem of what is commonly described as the *problem of multicollinearity*. In such a situation we should use only one set of the independent variable to make our estimate. In fact, adding a second variable, say X_2, that is correlated with the first variable, say X_1, distorts the values of the regression coefficients. Nevertheless, the prediction for the dependent variable can be made even when multicollinearity is present, but in such a situation enough care should be taken in selecting the independent variables to estimate a dependent variable so as to ensure that multi-collinearity is reduced to the minimum.

With more than one independent variable, we may make a difference between the collective effect of the two independent variables and the individual effect of each of them taken separately.

The collective effect is given by the coefficient of multiple correlations,

$R_{y \cdot x_1 x_2}$ defined as under:

$$R_{y \cdot x_1 x_2} = \sqrt{\frac{b_1 \sum Y_i X_{1i} - n\overline{Y}\,\overline{X}_1 + b_2 \sum Y_i X_{2i} - n\overline{Y}\,\overline{X}_2}{\sum Y_i^2 - n\overline{Y}^2}}$$

Alternatively, we can write

$$R_{y \cdot x_1 x_2} = \sqrt{\frac{b_1 \sum x_{1i} y_i + b_2 \sum x_{2i} y_i}{\sum Y_i^2}}$$

where

$$x_{1i} = (X_{1i} - \overline{X}_1)$$

$$x_{2i} = (X_{2i} - \overline{X}_2)$$

$$y_i = (Y_i - \overline{Y})$$

and b_1 and b_2 are the regression coefficients.

EXAMPLE:

A paint manufacturing company has researched that its sale mainly depends upon the advertisements in electronic media and sales promotion activities. Company increased its budget on above two heads considerably. Now company is interested to see the impact of these two initiatives on the sales. You are given with a random sample of the sales for 15 days. You are required to develop a multiple regression model and predict the impact of above two techniques on sales.

Also predict the sales when expenditure on advertising in electronic media is Rs120 thousand and on promotion activities is Rs 60 thousands.

Days	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Sales (‘000)	1000	1125	930	710	930	1000	1200	950	1110	1710	1590	1090	995	1400	1450
Adv (‘000)	9	9	18	18	27	27	38	38	52	52	58	58	65	65	75
Promotion (‘000	28	28	55	55	10	10	28	28	39	39	45	45	53	53	62

Solution:

The proposed multiple regression model is as below:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Where

Y = sales of company

X_1= Expenditure on advertisement in electronic media

X_2 = Expenditure on promotion activities

We can get the values of b_0, b_1, and b_2 (sample coefficients) by solving above three linear equations. These values provide the estimates of β_0, β_1, and β_2 (population parameters).

$$\sum Y_i = na + b_1 \sum X_{1i} + b_2 \sum X_{2i}$$
$$\sum X_{1i} Y_i = a \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i}$$
$$\sum X_{2i} Y_i = a \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2$$

By solving above three equations we can get value of b_0, b_1, and b_2 that are equivalent to β_0, β_1, and β_2.

	Unstandardized Coefficients	Std. Error	Standardized Coefficients Beta	t value	Sig.

(Constant)	833.9565767	127.8487		6.52299578	0.0000193
Adv	7.685798603	2.795796	0.606317935	2.749055591	0.016568
Promotion	-1.733	4.54487	-0.10034	0.465437	0.102409

β_0 = 833.956; β_1 = 7.685799; β_2 = -1.733

Now, we will test the significance of above calculated coefficient so that we can take a decision to include or exclude them in the proposed model. We can do it by using t-test as described in this module. Here, we consider the results obtained directly from SPSS output.

In above table we can see that value of **β_0** and **β_1** is found significant ($p< 0.05$). Value of **β_2** is found insignificant ($p > 0.05$). Therefore, we will not include **β_2** in our model. Therefore, now our model will be as:

$Y= \beta_0 + \beta_1X_1+ \beta_2X_2$

Sales = 833.956 + 7.68579 (exp on advertisement)……………………….(1)

The above equation gives us the estimated value of sales for any given value of expenditure on advertisements.

Y	**(y-ȳ)**	**(y-ȳ)²**	**X1**	**X2**	**Ŷ**	**(Ŷ-ȳ)**	**(Ŷ-ȳ)²**	**(Y-Ŷ)**	**(Y-Ŷ)²**
1000	-146	21316	9	28	903.125	-242.875	58988.2656	96.875	9384.76563
1125	-21	441	9	28	903.125	-242.875	58988.2656	221.875	49228.5156
930	-216	46656	18	55	972.29	-173.71	30175.1641	-42.29	1788.4441
710	-436	190096	18	55	972.29	-173.71	30175.1641	-262.29	68796.0441
930	-216	46656	27	10	1041.46	-104.545	10929.657	-111.455	12422.217
1000	-146	21316	27	10	1041.46	-104.545	10929.657	-41.455	1718.51702
1200	54	2916	38	28	1125.99	-20.01	400.4001	74.01	5477.4801
950	-196	3841	38	28	1125	-20.01	400.40	-175.99	30972.4801

		6			.99		01		
1110	-36	1296	52	39	1233.58	87.58	7670.2564	-123.58	15272.0164
1710	564	318096	52	39	1233.58	87.58	7670.2564	476.42	226976.016
1590	444	197136	58	45	1279.69	133.69	17873.0161	310.31	96292.2961
1090	-56	3136	58	45	1279.69	133.69	17873.0161	-189.69	35982.2961
995	-151	22801	65	53	1333.49	187.485	35150.6252	-338.485	114572.095
1400	254	64516	65	53	1333.49	187.485	35150.6252	66.515	4424.24522
1450	304	92416	75	62	1410.34	264.335	69872.9922	39.665	1573.31223
1146.0		**1067210.0**		**1146.0**		**392247.8**		**674880.7**	

The above table contains the detailed information about various statistics.

Sum of square of total variations (SST) = **1067210.0**

Sum of square of Regression (SSR) = **392247.8**

Sum of square of error (SSE) = **674880.7**

Multiple Coefficient of determination (R^2)

$R^2 = SSR / SST$

$= Explained\ Variations / Total\ Variations$

= 392247.8/ 1067210.0 = 0.40

It means proposed model explains only 40 percent variations in sales. The model is not good fit for the estimation purpose.

Standard Error of Estimation (Se):

$$S_e = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n-k-1}}$$

$$= \sqrt{\frac{674880.7}{15-2-1}} = \sqrt{\frac{674880.7}{12}} = 227.846$$

Calculation of F-statistics:

F-statistics is calculated to see that dependent variable is correlated with at least one independent variable. If value of F is found to be significant, it means that our model is significant in estimating the value of Y for the given values of Xi.

$$F = \frac{MSR}{MSE} = \frac{\text{Mean of sum of squares of explained variations}}{\text{Mean of sum osquares of unexplained variations}}$$

$$MSR = \frac{SSR}{k} \text{ and } MSE = \frac{SSE}{n-k-1}$$

Where

n = Number of observations

k = Number of independent variables

$$MSR = \frac{674880.7}{2} = 337440$$

$$MSE = \frac{674880.7}{15-2-1} = \frac{674880.7}{12} = 56240.1$$

$$F = \frac{MSR}{MSE} = \frac{337440}{56240.1} = 6 \quad (p < 0.05)$$

Thus, value of F is found to be significant, it means that our model may be used for estimation purpose.

Estimation of Sales (Y) when X_1 is Rs120 thousand and X_2 is Rs 60 thousands:

From equation (1)

Sales = 833.956 + 7.68579 (exp on advertisement)……………………..(1)

Sales = 833.956 + 7.68579 (120)

Sales = Rs 1756.251 thousands

DISCRIMINANT ANALYSIS:

Through Discriminant analysis technique, researcher may classify individuals or objects into one of two or more mutually exclusive and exhaustive groups on the basis of a set of independent variables. Discriminant analysis requires interval independent variables and a nominal dependent variable. For example, suppose that brand preference (say brand *x* or *y)* is the dependent variable of interest and its relationship to an individual's income, age, education, etc. is being investigated, and then we should use the technique of Discriminant analysis. Regression analysis in such a situation is not suitable because the dependent variable is, not intervally scaled. Thus Discriminant analysis is considered an appropriate technique when the single dependent variable happens to be non-metric and is to be classified into two or more groups, depending upon its relationship with several independent variables which all happen to be metric. The objective in Discriminant analysis happens to be to predict an object's likelihood of belonging to a particular group based on several independent variables. In case we classify the dependent variable in more than two groups, then we use the name multiple Discriminant analysis; but in case only two groups are to be formed, we simply use the term Discriminant analysis.

We may briefly refer to the technical aspects relating to Discriminant analysis.

(i) There happens to be a simple scoring system that assigns a score to each individual or object. This score is a weighted average of the individual's numerical values of his independent variables. On the basis of this score, the individual is assigned to the 'most likely' category. For example, an individual is 20 years old, has an annual income of Rs 12,000, and has 10 years of formal education. Let b_1, b_2, and b_3 be the weights attached to the independent variables of age, income and education respectively. The individual's score (z), assuming linear score, would be:

$$z = b_1 (20) + b_2 (12000) + b_3 (10)$$

This numerical value of z can then be transformed into the probability that the individual is an early user, a late user or a non-user of the newly marketed consumer product (here we are making three categories viz. early user, late user or a non-user).

(ii) The numerical values and signs of the *b*'s indicate the importance of the independent variables in their ability to discriminate among the different classes of individuals. Thus, through the Discriminant analysis, the researcher can as well determine which independent variables are most useful in predicting whether the respondent is to be put into one group or the other. In other words, Discriminant analysis reveals which specific variables in the profile account for the largest proportion of inter-group differences.

(iii) In case only two groups of the individuals are to be formed on the basis of several independent variables, we can then have a model like this

$z_i = b_0 + b_1X_{1i} + b_2X_{2i} + ... + b_nX_{ni}$

where X_{ji} = the *i*th individual's value of the *j*th independent variable;

b_j = the Discriminant coefficient of the *j*th variable;

z_i = the *i*th individual's Discriminant score;

$z_{crit.}$ = the critical value for the Discriminant score.

The classification procedure in such a case would be

If $z_i > z_{crit.}$, classify individual *i* as belonging to Group I

If $z_i < z_{crit}$, classify individual *i* as belonging to Group II.

When *n* (the number of independent variables) is equal to 2, we have a straight line classification boundary. Every individual on one side of the line is classified as Group I and on the other side, everyone is classified as belonging to Group II. When $n = 3$, the classification boundary is a two-dimensional plane in 3 space and in general the classification boundary is an $n - 1$ dimensional hyper-plane in *n* space.

(iv) In *n*-group Discriminant analysis, a Discriminant function is formed for each pair of groups. If there are 6 groups to be formed, we would have 6(6 – 1)/2 = 15 pairs of groups, and hence 15 Discriminant functions. The *b* values for each function tell which variables are important for discriminating between particular pairs of groups. The *z* score for each Discriminant function tells in which of these two groups the individual is more likely to belong. Then use is made of the transitivity of the relation "more likely than". For example, if group II is more likely than group I and group III is more likely than group II, then group III is also more likely than group I. This way all necessary comparisons are made and the individual is assigned to the most likely of all the groups. Thus, the multiple-group Discriminant analysis is just like the two-group Discriminant analysis for the multiple groups are simply examined two at a time.

(v) For judging the statistical significance between two groups, we work out the Mahalanobis statistic, D^2, which happens to be a generalized distance between two groups, where each group is characterized by the same set of *n* variables and where it is assumed that variance covariance structure is identical for both groups. It is worked out thus:

$$D^2 = (U_1 - U_2)\, v^{-1}\, (U_1 - U_2)'$$

where U_1 = the mean vector for group I

U_2 = the mean vector for group II

v = the common variance matrix

By transformation procedure, this D^2 statistic becomes an F statistic which can be used to see if the two groups are statistically different from each other.

From all this, we can conclude that the Discriminant analysis provides a predictive equation, measures the relative importance of each variable and is also a measure of the ability of the equation to predict actual class-groups (two or more) concerning the dependent variable.

Research report

Mostly, research work is presented in a written form. The practical utility of research study depends heavily on the way it is presented to those who are expected to act on the basis of research findings. Research report is a written document containing key aspects of research project.

Research report is a medium to communicate research work with relevant people. It is also a good source of preservation of research work for the future reference.

In simple terms, research report is the systematic, articulate, and orderly presentation of research work in a written form.

Types of research report:

The form and structure of the research report might change according to the purpose for which it has been designed. Based on the size of the report, it is possible to divide the report into the following types:

I. **Brief Reports**

These kinds of reports are not formally structured and are generally short, sometimes not running more than four to five pages. The information provided is of a limited scope and is prepared either for immediate consumption or as a prelude to the formal structured report that would subsequently follow. These reports could be designed in several ways.

- ***Working papers*** or ***basic reports*** are written for the purpose of collating the process carried out in terms of scope and framework of the study, the methodology followed and instrument designed. The results and findings would also be recorded here. However, the interpretation of the findings and study background might be missing, as the focus is more on the present study rather than past literature. These reports are significant as they serve as a reference point when writing the final report or when the researcher wants to revisit the detailed steps followed in collecting the study-related information.
- ***Survey reports*** might or might not have an academic orientation. The focus here is to present findings in easy-to-comprehend format that

includes figures and tables. The reader can then study the patterns in findings to arrive at appropriate conclusions, essential for resolving the business dilemma. The advantage of these reports is that they are simple and easy to understand and present the findings in a clear and usable format.

II. **Detailed Reports**

These are more formal and pedantic in their structure and are essentially academic, technical or business reports. Sometimes, the researcher may prepare both kinds—for an academic as well as for a business purpose. The language, presentation and format of the two kinds of reports would be vastly different as they would need to be prepared for the understanding of the reader's capabilities and intentions.

III. **Technical Reports**

These are major documents and would include all elements of the basic report, as well as the interpretations and conclusions, as related to the obtained results. This would have a complete problem background and any additional past data/records that are essential for comprehending and interpreting the present study output. All sources of data, sampling plan, data collection instrument(s), data analysis outputs would be formally and sequentially documented.

IV. **Business Reports**

These reports would not have the technical rigour and details of the technical report and would be in the language and include conclusions as understood and required by the business manager. The tables, figures and numbers of the first report would now be pictorially shown as bars and graphs and the reporting tone would be more in business terms rather than in conceptual or theoretical terms. If needed, the tabular data might be attached in the appendix.

Significance of research report:

- The research report fulfills the historical task of serving as a concrete proof of the study that was undertaken. This serves the purpose of providing a framework for any work that can be conducted in the same or related areas.

- It is the complete detailed report of the research study undertaken by the researcher, thus it needs to be presented in a comprehensive and objective manner. This is a one-way communication of the researcher's study and analysis to the reader/ manager, and thus needs to be all-inclusive and yet neutral in its reporting.
- For academic purpose, the recorded document presents a knowledge base on the topic under study and for the business manager seeking help in taking more informed decisions; the report provides the necessary guidance for taking appropriate action.
- As the report documents all the steps followed and the analysis carried out, it also serves to authenticate the quality of the work carried out and establishes the strength of the findings obtained.

Thus, effective recording and communicating of the results of the study becomes an extremely critical step of the research process. Based on the nature of the research study and the researcher's orientation, the report can take different forms.

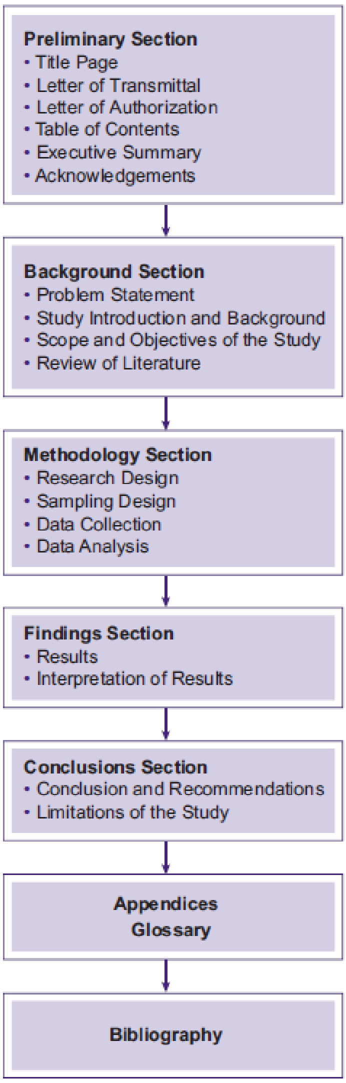
Structure of research report:

Whatever the type of report, the reporting and dissemination of the study and its findings require a structured format and by and large, the process is standardized. As stated above, the major difference amongst the types of reports is that all the elements that essentially constitute a research report would be present only in a detailed technical report. In the management report, the information on the sampling techniques follows the research intention, and the questionnaire design details need not be reported. The review of past literature would be perfunctory in the management report; however, they would be detailed and accompanied with the bibliography in the technical report. Usage of theoretical and technical jargon would be higher in the technical report and visual presentation of data would be higher in the management report.

The process of report formulation and presentation is presented in the below Figure. As can be observed, the preliminary section includes the rudimentary parts, for example the title page, followed by the letter of authorization, acknowledgements, executive summary and the table of contents. Then come the background section, which includes the problem statement, introduction, study background, scope and objectives of the study and the review of literature (depends on the purpose). This is followed by the methodology section, which, as stated earlier, is again specific to the technical report. This is followed by the findings section and then come the conclusions. The technical report would have a detailed bibliography at the end.

In the management report, the sequencing of the report might be reversed to suit the needs of the decision-maker, as here the reader needs to review and absorb the findings. Thus, instead of simply summarizing the statistical results, the findings need to be presented in such a way that they can be used directly as inputs for decision-making. Thus, the last section would be presented immediately after the study objectives and a short reporting on methodology could be presented in the appendix.

Thus, the entire research project needs to be recorded either as a single written report or into several reports, depending on the need of the readers. The researcher would need to assist the business manager in deciphering the report, executing the findings, and in case of need, to revise the report to suit the specific actionable requirements of the manager.



Preliminary Section

This section mainly consists of identification information for the study conducted. It has the following individual elements:

Title page: This includes classification data about:

• The target audience, or the intended reader of the report.

• The report author(s), including their name, affiliation and address.

• The title of the study presented in a manner to clearly indicate the study variables; the relationship or status of the variables studied and the population to which the results apply. The title should be crisp and indicative of the nature of the project, as illustrated in the following examples.

- ✓ Comparative analysis of BPO workers and schoolteachers with reference to their work–life balance.
- ✓ Segmentation analysis of luxury apartment buyers in the National Capital Region (NCR).
- ✓ An assessment of behavioural factors impacting consumer financial investment decisions.

Letter of transmittal: This is the letter that goes alongside the formalized copy of the final report. It broadly refers to the purpose behind the study. The tone in this note can be slightly informal and indicative of the rapport between the client-reader and the researcher. The letter broadly refers to three issues. It indicates the term of the study or objectives; next it goes on to broadly give an indication of the process carried out to conduct the study and the implications of the findings. The conclusions generally are indicative of the researcher’s interest/learning from the study and in some cases may be laying the foundation for future research opportunities.

Letter of authorization: Sometimes the letter of authorization may be redundant as indications of the formal approval for conducting the study might be included in the letter of transmittal. The author of this letter is the business manager or corporate representative who formally gives the permission for executing the project. The tone of this letter, unlike the above document, is very precise and formal, leaving no room for speculation or interpretation.

As explained, this letter is not critical to submission, in case reference to the same has been made in the transmittal letter. However, in case it is to be included in the report, it is advisable to reproduce the exact prototype of the original letter.

Table of contents: All reports should have a section that clearly indicates the division of the report based on the formal areas of the study as indicated in the research structure. The major divisions and subdivisions of the study, along with their starting page numbers, should be presented. The subheadings and the smaller sections of a topic need not be indicated here as then the presentation of the content seems cluttered.

Once the major sections of the report are listed, the list of tables come next, followed by the list of figures and graphs, exhibits (if any) and finally the list of appendices.

Executive summary: This is the last and the most critical element of the preliminary section. The summary of the entire report, starting from the scope and objectives of the study to the methodology employed and the results obtained, have to be presented in a brief and concise manner. In case the research requirement was to provide recommended changes based on the findings, it is advisable to provide short pointers here. Interestingly, it has been observed that in most instances the business managers read only the executive summary in its complete detail and most often just glance through the rest of the report. Thus, it becomes extremely critical to present a Gestaltan view of the entire report in a suitable condensed form. The executive summary essentially can be divided into four or five sections. It begins with the study background, scope and objectives of the study, followed by the execution, including the sample details and methodology of the study. Next comes the findings and results obtained. The fourth section covers the conclusions which are more or less based on the opinion of the researcher. Finally, as stated earlier, in case the study objectives necessitate implications, the last section would include recommendations and suggestions.

Acknowledgements: A small note acknowledging the contribution of the respondents, the corporate and the experts who provided inputs for accomplishing the study is to be included here.

Though the executive summary comes before the main body of the report, it is always prepared after the entire report has been finalized and is ready in its final form. The length of this section is one or two pages only and the researcher needs to effectively present the most significant parts of the study in a succinct form. It has been observed that the executive summary is a standalone document that is often circulated independently to the interested managers who might be directly or indirectly related to the study.

Main Report

This is the most significant and academically robust part of the report. The sections of this division follow the essential pattern of a typical research study.

Problem definition: This section begins with the formal definition of the research problem. The problem statement is the research intention and is more or less similar to what was stated earlier as the title of the research study.

Study background: Study background presents details of the preliminary conceptualization of the management decision problem and all the groundwork done in terms of secondary data analysis, industry experts' perspectives and any other earlier reporting of similar approaches undertaken. Thus, essentially, the section begins by presenting the decision-makers' problem and then moves on to a description of the theoretical and contemporary market data that laid the foundation that guided the research.

In case the study is an academic research, there is a separate section devoted to the review of related literature, which presents a detailed reporting of work done on the same or related topic of interest.

Study scope and objectives: The logical arguments then conclude in the form of definite statements related to the purpose of the study. A clear definition of the scope and objective of the study is presented usually after the study background; in case the study is causal in nature, the formulated hypotheses are presented here as well.

Methodology of research: This section would not be sequentially placed here, for short reports or for a business report. In such reports, a short description of the methodology followed would be documented in the appendix. However, for a technical and academic report, this is a significant and primary contribution of the research study. The section would essentially have five to six sections specifying the details of how the research was conducted. These would essentially be:

- *Research framework or design:* The variables and concepts being investigated are clearly defined, with a clear reference to the relationship being studied. The justification for using a particular design has to be presented in a sequential and step-wise manner enlisting the experimental and control conditions, in case of a causal study. The researcher must take care to keep the technical details of the execution in the appendix and present the execution details in simple language, in the main body.
- *Sampling design:* The entire sampling plan in terms of the population being studied, along with the reasons for collecting the study-related information from the given group is given here. The execution details, in terms of sample size calculations, sampling frame considered and field work details can be recorded in the appendix rather than in the main body of the report. However, the sample profile and identification details are included in the main section. As stated earlier, the report needs to be reader-friendly, and too much technical information might not be required by the decision-maker.
- *Data collection methods:* In this section, the researcher should clearly list the information needed for the study as drawn from the study objectives stated earlier. The secondary data sources considered and the primary instrument designed for the specific study are discussed here. However, the final draft of the measuring instrument can be included in the appendix, which includes the execution details in terms of how the information was collected; how the open ended or opinion-based questions were handled; and how irregularities were handled and accounted for in the study. These and similar information enable a clear insight into the standardization of procedures maintained.
- *Data analysis:* Here, the researcher again needs to revisit the research objectives and the study design in order to justify the analytical tools and techniques used in the study. The assumptions and constraints of the analysis need to be explained here in simple, non-technical terms. There is no need to give a detailed description of the statistical calculations here.
- *Study results and findings:* This is the most critical chapter of the report and requires special care; it is probably also one of the longest chapters in

the document. The researcher could, thus, consider either breaking this into subchapters or at least clear subheadings.

Interpretations of Results and Suggested Recommendations

The section study results and findings, i.e., the main report, presents a bird's eye view of the information as it exists in a summarized and numerical form. This kind of information might become difficult to understand and convert into actionable steps, thus the real skill of the researcher lies in simplifying the data in a reader-friendly language. Here, it is recommended that this section should be more analytical and opinion based. The results could be supported by the data that was presented earlier, for example, industry forecasts or the expert opinion. In case the report had an earlier section on literature review, the researcher could demonstrate the similarity of findings with past studies done on the topic.

Limitations of the Study

The last in this section is a brief discussion of the problems encountered during the study and the constraints in terms of time, financial or human resources. There could also have been constraints in obtaining the required information, either because the data about the topic of interest has not been collected or because it is not readily available to all. These clear revelations about the drawbacks are thus kept in mind by the reader when analysing the results and the implications of the study.

End Notes

The final section of the report provides all the supportive material in the study. Some of the common details presented in this section are as follows:

Appendices: The appendix section follows the main body of the report and essentially consists of two kinds of information:

1. Secondary information like long articles or in case the study uses/is based on/ refers to some technical information that needs to be understood by the reader. Or long tables or articles or legal or policy documents.

2. Primary data that can be compressed and presented in the main body of the report. This includes: Original questionnaire, discussion guides, formula used

for the study, sample details, original data, long tables and graphs which can be described in statement form in the text.

Bibliography: This is an important part of the final section as it provides the complete details of the information sources and papers cited in a standardized format. It is recommended to follow the publication manuals from the American Psychological Association (APA) or the Harvard method of citation for preparing this section. In fact, with the advancement in computer technology the Microsoft office Word 2007 can automatically generate a bibliography based on any of these formats, based on the source information provided in the document.

The reporting content of the bibliography could also be in terms of:

• **Selected bibliography:** Selective references are cited in terms of relevance and reader requirement. Thus, the books or journals, that are technical and not really needed to understand the study outcomes are not reported.

• **Complete bibliography:** All the items that have been referred to, even when not cited in the text, are given here.

• **Annotated bibliography:** Along with the complete details of the cited work, some brief information about the nature of information sought from the article is given. This could run into three or four lines or a brief paragraph.

At this juncture we would like to refer to another method of citation that an author might wish to use during report writing. This could be in the form of a footnote. To explain the difference we would first like to explain what a typical footnote is:

Footnote: A typical footnote, as the name indicates, is part of the main report and comes at the bottom of a page or at the end of the main text. This could refer to a source that the author has referred to or it may be an explanation of a particular concept referred to in the text.

The referencing protocol of a footnote and bibliography is different. In a footnote, one gives the first name of the person first and the surname next. However, this order is reversed in the bibliography. Here we start first with the surname and then the first name. In a bibliography, we generally mention the page numbers of the article or the total pages in the book. However, in a footnote, the specific page from which the information is cited is mentioned. A bibliography is generally arranged alphabetically depending on the author's name, but in the footnote the reporting is based on the sequence in which they occur in the text.

Glossary of terms: In case there are specific terms and technical jargon used in the report, the researcher should consider putting a glossary in the form of a word list of terms used in the study. This section is usually the last section of the report.

Presentation of report (oral):

Once the final draft of the research report is prepared and documented, the last stage is sharing the findings and research implications with the client or interested audience. This is usually done orally and with the support of visual aids. The presentation that the researcher might be making could be detailed for his team members or for an academic audience. However, in case the presentation is for the client or for a business audience, brevity and focus of the presentation is critical. A thumb rule for this is not to go beyond 20 minutes with more time for question and answers and interactive discussion on the findings.

Regardless of the audience for the presentation, the most critical aspect of the presentation is two-fold:

(a) Who is the listener? What does he/she seek from the presentation?

(b) What is the core of the briefing—is it background, or methodology, key findings or decision directions that the findings are indicating?

Once the researcher is clear on this, he needs to need to focus on three key aspects:

Study background: This should be essentially 10–15 per cent of the entire presentation. It should explain the impetus behind the study as briefly and with suitable emphasis as possible.

Study findings: The major conclusions of the study need to be shared in simple words and with appropriate supportive visuals or material. The researcher must be able to demonstrate clearly the link between the study objectives and the findings.

Study implications: In case this was agreed upon between the researcher and the client or was specified as a study objective by the researcher, this section would be the last section of the presentation. The link between what was found and what is suggested must be clear to the audience. The researcher may vary the discussion time between the earlier section and this as 45 per cent each or 30–70 or 70–30, depending on the study objective, i.e., more findings or more implication oriented. As supportive material the researcher can make use of:

Handouts: These could be in the form of the primary questionnaire designed for the study or company brochures and other related secondary material. They should be distributed to the audience when the presenter is referring to them.

Slides: These are created today with the help of computer programmes. There are endless possibilities enhancing the material be presented and for engaging the listener. The designing and creation of the material requires considerable skill and care to ensure that the presentation style should be the supportive aid for an effective delivery and not a showcase of the computer graphics that the researcher is well versed with. Too much clutter and a random mix of text and graphics should be avoided. Animation of the data in synchronization with the vocal delivery makes the presentation more forceful.

Chalkboards and flipcharts: These are additional visual aids that could be kept as standby for the question-and-answer session when an idea might have to be highlighted or demonstrated in the response of some query raised by the listeners. However, use of these means during an active presentation should be avoided as they necessitate the presenter to be engaged with the medium at the cost of losing contact with the listener.

Video and audio tapes: Again, these are supportive materials that can be used to emphasize a point.