Chapter III

RESEARCH

METHODOLOGY
Chapter III
RESEARCH METHODOLOGY

3.1 INTRODUCTION
Research\textsuperscript{136} refers to search for knowledge. Redman and Mory define research as a 'systematized effort to gain new knowledge\textsuperscript{137}' considering this fact researcher made systematic efforts to gain insight of present research study.

The present research study is a systematic and objective process of gathering, recording, and analyzing data to gain new knowledge about the problem under study. Brief explanations of the steps taken by the researcher in the present research study were given below.

3.2 IDENTIFYING AND DEFINING PROBLEM
The researcher has identified and defined the problem by conducting secondary data study. This well defined problem gave the researcher a proper direction for carrying out investigation.

3.3 FORMULATING RESEARCH OBJECTIVES
After defining the problem clearly, researcher set the research objectives and stated them in formal research statements.


\textsuperscript{137} L.V. Redman and A.V.H. Mory, The Romance of Research 1923, P.10
3.4 FRAMING RESEARCH HYPOTHESES

After setting research objectives researcher developed research hypotheses related to problem under study. If a hypothesized relationship or prediction has to be tested by scientific methods, it is called research hypothesis. A research hypothesis is one that links an independent variable to a dependent variable. It should generally contain one dependent and one independent variable.

3.5 DEPENDENT AND INDEPENDENT VARIABLES

A variable is a concept that can take on different quantitative values like height, weight, age and so on. If a variable is dependent on the result of some other variable, it is then called a dependent variable\textsuperscript{138}. An independent variable\textsuperscript{139} is one that is not dependent on any other variable with reference to that particular study. For instance, height and weight are dependent on age, but age is not dependent on height and weight. Therefore, age is an independent variable, while weight and height are dependent variables.

3.6 RESEARCH DESIGN

A research design\textsuperscript{140} is a framework that provides specific details regarding the process to be followed in conducting the research. The researcher has to select the appropriate research design after considering factors like the research objectives, costs involved, time allotted, and availability of data sources.

Research Design can be defined as the plan and structure of enquiry, formulated in order to obtain answers to research

\textsuperscript{138} ICFAI, Quantitative Methods II, 2006
\textsuperscript{139} ICFAI, Quantitative Methods II, 2006
\textsuperscript{140} ICFAI, Quantitative Methods II, 2006
questions. It can be understood as that plan which gives the blueprint for collection, measurement and analysis of data.

**Need for Research Design:** A good research design facilitates the smooth flow of the entire research process. It should utilize minimum time, money and effort. If the research design is not properly developed, it would fail the entire research process and will not meet its purpose. Further a good research design will clearly describe the techniques to be used for selecting samples, collecting data, and other aspects that are essential for conducting research. Based on the research design, decisions are taken depending on certain crucial issues like the purpose of study, objectives of study, the type of data needed, the method adopted for obtaining the data and analyzing it. To conclude, the research design is a tentative plan of number of steps implied in the entire research process. It also guides the researcher to take every step in the research systematically.

Hence with due importance researcher made all the efforts to design the research systematically. This research design acted as a thorough guideline for the researcher throughout the entire study and kept researcher focused on the set research objective. Research design for the present study in detail is as follows.

### 3.7 OVERVIEW OF RESEARCH METHODOLOGY

The researcher has to select the research method considering factors like research objectives, costs involved, the time allotted, and availability of data sources. In the current study researcher used survey method and secondary data study method for the purpose of research investigation. Data collection was done in two stages viz. Pilot Study and the Main Study.
3.7.1 PILOT STUDY

Pilot study is also called as Pre-testing which refers to testing the questionnaire on a small sample of respondents selected on a convenient basis that is not too divergent from the actual respondents. This helps reveal errors or flaws like wrong order of questions, leading questions and awkward responses. The reactions and attitudes of the respondents should be observed during this phase as it gives the researcher a firsthand experience of the potential problems that can be expected from the questionnaire. The questionnaire was revised after considering the identified flaws and a final draft questionnaire is prepared for conducting the actual survey.

Thus researcher selected approximately 5% sample from each population and conducted pilot study to understand the effectiveness of the questionnaire developed and to understand the problems involved in data collection for main study. After pilot study necessary modifications were done in questionnaire for serving the purpose of research investigation in better manner. This helps the researcher to minimize potential errors that may gather during the main study.

3.7.2 MAIN STUDY

After pilot study researcher conducted main study. Primary Data as well as Secondary Data were collected for the purpose of main study. Details of sources of data collection were as follows:
3.7.3 PRIMARY DATA

Primary Data is the data collected directly from respondents using data collection methods like survey interviews, measurements, direct observation etc. To collect the primary data a research tool, structured and segmented questionnaire have prepared for following two target groups of respondents for research investigations, Use of 5-point Likert Scale is made to measure qualitative data.

1. A structured, segmented questionnaire for MBA Students of various management institutes affiliated to University of Pune in Pune City.
2. A structured, segmented questionnaire for HR Managers, Executives, Functional Managers who are directly or indirectly part of recruitment process of MBA fresher.

3.7.4 SECONDARY DATA

Secondary data is the data which is already available and has already been collected by some other organization or person for their use, and is generally made available to other researchers for free or at a concessional rate. This data is available through websites, trade associations, journals, books, etc. The present research approach is multidimensional. Here researcher at the outset collected and studied secondary data to develop the theoretical and conceptual framework of present study. So following dimensions were designed to gain complete insight of the study

1. Review of related research articles and other articles.
3. Review of HR Tools and Techniques used for Career Planning and Development Activities
4. Study of Career Planning and Development initiative taken by selected Top World Class Universities.
5. Study of Development Profile of Pune City

The above mentioned Secondary Data was collected by review of various books, magazine, Internet website, newspapers.

3.7.5 SAMPLE DESIGN

Population or Universe: Population\textsuperscript{141} refers to the aggregation of all the items from which samples can be drawn. Here in present Study we have two sets of Population viz. said MBA Students and Functional Heads of various MNC's, involved directly or indirectly in the process of recruitment of MBA Fresher.

Sample: A sample\textsuperscript{142} is an individual element or a group of elements selected from the total population. Sampling is an act, process, or technique of selecting a part of population that represents the characteristics of the entire population with the help of sampling techniques.

\textsuperscript{141} ICFAI, Quantitative Methods II, 2006
\textsuperscript{142} ICFAI, Quantitative Methods II, 2006
Out of above mentioned various sampling techniques researcher used Stratified Random sampling method to select sample from respective sets of populations. Stratified random sampling is done by grouping the members of the population into relatively homogeneous subgroups called ‘strata’ before sampling. Then random sampling is applied within each stratum independently.

**Determining Sample Size for problems involving means:**
The sample size for problems involving means can be determined by using the following formula\(^{143}\),

\[
\text{Sample Size, } n = \left( \frac{2ZS}{E} \right)^2
\]

Where, \(E = ZS \bar{x}\)  \(\bar{x}\) standard error of the mean (\(S \bar{x}\)) = \(\frac{S}{\sqrt{n}}\)

\(^{143}\) ICFAI, Quantitative Methods II, 2006
Z = standardized value corresponding to a confidence level
S = Sample Standard Deviation or an estimate of the population standard deviation.
E = acceptable level of error, plus or minus an error factor (the range is one-half of the total confidence interval)
In case of a small or finite population, a finite point correction factor is used \( \frac{\sqrt{N-n}}{N-1} \).

**Sample Design**: As there were two different groups of respondents we have two sets of population and thereby two sets of sample for study. Following are the details of sampling techniques used to determine sample size.

**Sample Design – I**

Sampling Method Adopted by researcher was Probability Sampling. Researcher used Stratified Random Sampling method for selecting the Sample No. 1. Gender wise strata were selected, Male: Female ratio was maintained approximately 60:40.

**Population I**: The details of the MBA Institutes affiliated University of Pune is as follows:

**Table 3.1 The Details of the MBA Institutes affiliated University of Pune**

<table>
<thead>
<tr>
<th>Area Details</th>
<th>Total Number of MBA Institutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pune Municipal Corporation (PMC)</td>
<td>103</td>
</tr>
<tr>
<td>Pimpri Chinchwad Municipal Corporation (PCMC)</td>
<td>49</td>
</tr>
<tr>
<td>Pune Rural</td>
<td>74</td>
</tr>
<tr>
<td>Ahmed Nagar</td>
<td>27</td>
</tr>
<tr>
<td>Nashik</td>
<td>33</td>
</tr>
<tr>
<td>Silvasa</td>
<td>01</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>287</strong></td>
</tr>
</tbody>
</table>
In Pune City (In PMC area) we have total 103 MBA Institutes as mentioned in above table. Out of these 16 (15% of total population 103) institutes were selected for survey purpose. Total Sanctioned Intake of the selected institute for the purpose of study was 2520. Hence Sample Size, 20 % total sanctioned intake of 2520 = 504.

**Sample Size I:** Thus for first category of respondents that is MBA students of 16 Management Institutes affiliated to University of Pune, in Pune City, **The Sample Size I determined was 504 respondents**, which comprises of about 60% male students and 40 % Female Students.

**Sample Design – II**

Sampling Method Adopted by researcher was Probability Sampling. Researcher used Stratified Random sampling method for selecting the Sample No. 2. This includes 30 % of sample of HR Personnel, 20 % Marketing and Sales, 15 % General Management, 10 % belongs to IT , Operation and Finance each, 5 % from International Business Management area.

**Population II:** Total Corporate Personnel viz. HR Managers, Executives, Functional Heads who were directly or indirectly involved in the recruitment process of MBA fresher. Total no. of Pune City based Multinational Corporations (MNC) was 253.

Out of total 253 MNC’s, 17 % MNC’s were selected randomly for data collection. Thus 42 (Forty Two) MNC’s were selected and data were collected from 126 respondents representing these MNC’s to meet the above mentioned proportion of functional areas.

**Sample Size II:** Thus for the second category of respondents which includes HR Managers, Functional Heads and Executives
from 42 corporate in Pune City who were directly and indirectly involved in recruitment process of MBA Graduates. **The Sample Size II determined was 126 respondents** in which 30% belongs to HR, 15% belongs to General Management Area, 10% belongs to Finance, IT and Operation and 5% belongs to International Business Specialization.

**3.7.6 STATISTICAL TOOLS USED FOR DATA ANALYSIS AND INTERPRETATION**

Primary data collected for two sets of respondents viz. MBA Students and Corporate Personnel was analyzed by using software like Microsoft Excel, SPSS 16 V. Statistical Tools like Tables, Bar Diagram, Pie Charts, Percentages, Averages and Ratios etc were used to present and analyze the collected data. The researcher used bar diagram and Pie Charts, Line chart to show the trends at a glimpse. Further trends and behaviors of the variables were interpreted. For Testing of Hypotheses One Sample ‘Z’ Test and ‘Chi Square Test’ were applied. Details of data analysis and interpretations of present research study were discussed in detail in next chapter. ANOVA and Multiple Regression was used understand the trends of variables.

**3.7.7 DATA VALIDITY AND RELIABILITY**

**Validity:** Some of the important validity test generally considered includes content, convergent, discriminate and criterion related validity. According to Rungtusanatham\(^ {144} \) the content validity of a construct can be defined as the degree to which the measure spans the domain of the constructs. For the present study, the content

validity of the instrument was ensured as the career planning dimension were identified from the literature and thoroughly reviewed by the researcher.

External validity refers to how generalizable finding are across time, settings and individuals (Scandura and Williams)\textsuperscript{145}. The use of variety of methods may result in higher external validity. From this point of view, both review of various spectrums of career planning and counseling and questionnaire were contributed to a more robust and generalizable set of finding. With consideration given to the purpose of this study, it was believed that chosen sample and method are applicable. Internal validity can be described as the degree to which the researcher studies and measure what is intended (Merriam, 1994)\textsuperscript{146}. To improve the internal validity the study, additional discussions with professionals, observations were considered in the present research.

**Reliability:** The role of reliability is to minimize errors and biases in the study (Yin, 1994)\textsuperscript{147}. To enhance the reliability of the work it had been tried to describe the research methodology and strategy in such way that possible error, previously undetected, can be detected by the reader of this thesis. Regarding biases, there is always a problem with 'reactivity' the reaction on the part of those being investigated to the investigator and his or her research instrument (Bryman, 1988)\textsuperscript{148}. The consequences of reactivity are that people's behavior or view (Silverman, 2000)\textsuperscript{149}. The reliability of items was assessed by computing the coefficient alpha (Cronbach, 1951), that measure the

\textsuperscript{147} Yin, R.K. (1994).
internal consistency of the items. For a measure to be acceptable, coefficient alpha should be above 0.7 (Nunnally, 1978). Cronbach’s alpha can be written as a function of the number of test items and the average inter-correlation among the items. For conceptual understanding following is the formula for the standardized Cronbach’s alpha:

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N - 1) \cdot \bar{c}}$$

Where $N$ = Number of Items 
$\bar{C}$ = Average inter-item covariance among the items and 
$\bar{V}$ = Average variance.

Here it is observed that if number of items were increased Cronbach alpha was increased. Additionally, if the average inter item correlation is low, Cronbach alpha will be low. As the average inter item correlation increases, Cronbach alpha increases as well. To check the reliability of data researcher has used Cronbach alpha reliability coefficient. It measures the reliability of responses. It determines whether the responses are homogeneous or not by using various correlation coefficient. Thus Cronbach alpha reliability coefficient measures the reliability of questionnaire. Researcher determined the Cronbach alpha reliability coefficient by using SPSS 16 and value of Cronbach alpha reliability coefficients were 0.81 and 0.84 for questionnaire I and II respectively.

### 3.7.8 Hypotheses Testing

It is used to test the significance of a population parameter on the basis of sample. In testing of hypothesis, we make assumptions

---

151 ICFAI, Quantitative Methods II, 2006
regarding the value of population parameter. Then we collect sample data and arrive at the sample statistics. To test the validity of our assumption, we determine the difference between the population parameter and sample statistics. If the difference is small, then there is greater chance of the assumption being correct and if the difference is large then there is less chance of the assumption being correct. The purpose of hypothesis testing is not to question the computed value of the sample statistic but to make a judgment about the difference between that sample statistics and population parameter.

In the present study there were five research hypotheses. Researcher used One Sample Z Test and Chi-Square Test to test said hypotheses. Following were standard steps for hypothesis testing. Specify null and alternative hypothesis, Select significance level, Compute test statistics based on an appropriate probability distribution, Locate the critical region, Describe the result and the statistical conclusion.

Test Statistics = Sample Statistics – Hypothesized Parameter

\[
\text{Standard Error of Statistics}
\]

Once we have specified the null hypothesis and the alternative hypothesis, we need to calculate the test statistics. The choice of test statistics differs with the size of the sample.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>When the Population Standard Deviation is Known</th>
<th>When the Population Standard Deviation is Not Known</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size Greater than 30</td>
<td>Normal Distribution, z Table</td>
<td>Normal Distribution, z Table</td>
</tr>
<tr>
<td>Sample size Less than or equal to 30</td>
<td>Normal Distribution, z Table</td>
<td>t Distribution, t Table</td>
</tr>
</tbody>
</table>
In the present research one sample Z test is used, as it is the case of large sample. Following is the formula for computing value of Z. 

One sample z test is used to check specified value of arithmetic mean in same sample (one sample).

\[
Z = \frac{X - \mu}{(s / \sqrt{n})}
\]

Where X = Arithmetic Mean  
\(\mu\) = Estimated Mean of Population  
S = Standard Deviation  
n = Sample Size

In the present study two tailed test is applied. In two tailed tests of a hypothesis will reject the null hypothesis if the sample mean is significantly higher than or lower than the hypothesized population mean.

**Concept of Degree of Freedom:** Degrees of Freedom = Number of observations - Number of independent constraints. The degrees of freedom can be calculated as follows\(^{152}\):

For Two Samples \(df = (\text{Number of Rows - 1})(\text{Number of Columns - 1})\)

For One Sample \(df = \text{Sample Size} - 1\)

Degrees of freedom (df) refers to the number of values that are free to vary after restriction has been placed on the data. Degrees of freedom (df) determines shape of the distribution.

**Level of Significance:** In hypothesis testing, the significance level is the criteria used for rejecting the null hypothesis\(^ {153}\).

**Chi-Square Test:** In the test of significance of mean we are comparing the mean of one sample with the hypothesized

---

\(^{152}\) ICFAI, Quantitative Methods II, 2006  
\(^{153}\) ICFAI, Quantitative Methods II, 2006
population mean. In the test of significance of difference between two means, we are comparing the means of two samples. In chi-square test, we can check the equality of more than two population parameters like proportions, means.

Chi-Square test is one of the most powerful tests used in inference statistics. Chi-square test has various applications some are as follows. The first application is to test the difference of proportion in two or more samples. The second application is to test the proportion of one characteristic in one sample and last but not the least the one of the most important application is to test goodness of fit to find out association between two attributes. Here in present study researcher used to test the proportion of one characteristic in one sample. If we classify a population into several categories with respect to two attributes we can then use a chi-square test to determine whether the two attributes are independent of each other. The observed frequencies are the actual frequencies, whereas the expected frequencies or the theoretical frequencies are the estimated frequencies.

Following steps were followed to solve problems involving chi-square statistics. State the null and alternative hypothesis, List the observed frequencies, Calculate Expected Frequencies, Calculate the difference between the corresponding and observed and expected frequencies, Express the square of the difference as fraction of corresponding expected frequency, Add all the fractions obtained in above step no.5, Compare the value with the appropriate $X^2$ value from the table, Accept the null hypothesis if the value thus computed for the given degrees of freedom and levels of significance is lesser than the critical (tabulated value), otherwise reject it.
Formula for chi-square Test:

\[
\chi^2 = \frac{k(O_i - E_i)}{\sum_{i=1}^k E_i}
\]

Where

- \(O\) = Observed Frequency
- \(E\) = Expected Frequency

3.7.9 CORRELATION AND REGRESSION

Correlation\textsuperscript{154} is a statistical tool that helps to measure and analyze the degree or extent to which two or more variables fluctuate with reference to one another. In other words, if the change in one variable affects a change in the other variable, the variables are said to be correlated. Thus, correlation denotes the inter-dependence among the variables. For correlating two phenomena, it is essential that the two phenomena should have a cause-effect relationship and if such relationship does not exist, then the two phenomena cannot be correlated.

The degree of relationship is expressed by a coefficient which ranges from -1 to +1. The direction of change is indicated by a negative or positive sign. Correlation is of four types as listed: Positive and Negative, Simple and Multiple, Partial and Total, Linear and Non-linear. Important methods of studying correlation are Scatter diagram, Karl Pearson’s coefficient of correlation, and Spearman’s Rank Correlation Coefficient. Researcher used Spearman’s Rank Correlation Coefficient to study correlation in the present study. In the present study correlation is studied between independent variables career anchors and dependent variable career options.

\textsuperscript{154} ICFAI, Quantitative Methods II, 2006
**Spearman’s Rank Correlation Coefficient:** This method is especially useful when quantitative measure of certain factors involving quantity or quality cannot be fixed. Cases involving the evaluation of leadership skills and ability of an individual or the judgment of female beauty and the like may make use of this method. Spearman's Rank correlation coefficient is given by:

\[ R = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \]

Where,
- \( R \) = Rank coefficient of correlation
- \( D \) = Difference of ranks between paired item in two series
- \( N \) = Total number of observations

**Interpretation of Rank Correlation Coefficient:** The value of rank correlation coefficient, \( R \) ranges from -1 to +1. If \( R = +1 \), then there is complete agreement in the order of the ranks and the ranks are in the same direction. If \( R = -1 \), then there is complete agreement in the order of the ranks and the ranks are in the opposite direction. If \( R = 0 \), then there is no correlation.

**Regression Analysis:** In the present research study regression analysis\(^{155}\) is used to study the relationship between two variables viz. Independent variable career anchors and dependent variable career options. As this study is real life situations, a phenomenon is affected by the more than one variable. To study such phenomenon, multiple regressions were used. In multiple regressions the linear relationship between the dependent variable and more than one independent variable can be studied. In general, the multiple regression equation is given by:

---

\(^{155}\) ICFAI, Quantitative Methods II, 2006
\[ \hat{y} = A + B_1X_1 + B_2X_2 + B_3X_3 \ldots \ldots + B_nX_n \]

\[ \hat{y} \] = Estimated value corresponding to the dependent variable

A = Y-intercept

Where,

X₁, X₂, X₃ \ldots \ldots Xₙ Independent variables

B₁, B₂ \ldots \ldots, Bₙ Slopes associated with X₁, X₂, \ldots \ldots Xₙ

**Standard Error in Multiple Regressions:** The standard error of the estimate for a regression equation is given by

\[ Se = \sqrt{\frac{\sum(Y - \hat{y})^2}{n - k - 1}} \]

Where,

Y = Sample value of the dependent variable

Ye = Corresponding estimate value obtained by using the Regression Equation.

n = Number of observations in the sample

k = Number of independent variables

For the regression equation with two independent variables, an alternative formula mainly utilizing data already collected for ascertaining the regression equation is:

\[ Se = \sqrt{\frac{\sum(Y^2 - aY - b_1 \Sigma X_1Y - b_2 \Sigma X_2Y)}{(n - k - 1)}} \]

The estimates made by multiple regressions can be improved if the number of the observations used for estimation is increased.

**Coefficient of Multiple Determinations (R²):** In multiple correlations, we measure the strength of the relationship among three variables using the co-efficient of multiple determinations, R² or its square root, R (the coefficient of multiple correlations). This coefficient of multiple determinations is also the proportion
of the total variation of \( Y \) that is explained by the regression plane.

\[
R^2 = \frac{\text{Variation explained by the model}}{\text{Total variation}}
\]

\[
\begin{align*}
R^2 &= \frac{\sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\text{ESS}}{\text{TSS}} \\
\text{TSS} &= \text{(Total sum of Squares)} = \sum (Y - \bar{Y})^2 \\
\text{RSS} &= \text{(Regression sum of Squares)} = \sum (Ye - \bar{Y})^2 \\
\text{ESS} &= \text{(Error sum of Squares)} = \sum (Y - Ye)^2 \\
\text{TSS} &= \text{RSS} + \text{ESS}
\end{align*}
\]

The coefficient of multiple determinations takes values in the range 0 to 1. Thus, \( R^2 \) explains the percentage of variation in \( Y \) which is explained by the multiple regression model with \( X_1 \) and \( X_2 \) as the independent variables. For example, if \( R^2 = 0.75 \), it means that 75% of the variation in the dependent variable can be explained by the multiple regression equation in terms of independent variables and rest 25% is explained by other factors. Details of Hypotheses Testing along with test statistics and details of Multiple Regression analysis is given in the next chapter.