CHAPTER – 3
RESEARCH METHODOLOGY

3.1 Introduction

3.1 Methodology

3.1.1 Research Design

3.1.2 Research Framework Design

3.1.3 Research Instrument

3.1.4 Validity of Questionnaire

3.1.5 Statistical Measurement

3.1.6 Target Population and Sample Size

3.1.7 Survey Procedure

3.2 Data Analysis

3.2.1 Chi Square

3.2.2 ANOVA

3.2.3 Discriminate Function

3.2.4 Factor Analysis

3.2.5 Garret Ranking

3.2.6 Correlation and Path Analysis

3.2.7 Regression Analysis

3.3 Conclusion
3.0  INTRODUCTION

This chapter discusses a detailed account of the research methodology, including explanation about survey, research design, research instrument, and various statistical tools applied in this research.

3.1  METHODOLOGY

3.1.1  Research Design

This study employed a casual research design using a customer survey, aiming at developing and testing a conceptual customer perception and satisfaction model, in which hypothesis was advanced and tested to confirm casual relationships. A self administered questionnaire was developed to collect the data. The duration of data collection of the study extended from August 2008 to May 2009.

3.1.2  Research Framework Design

Figure–3.1 displayed the research framework for the study. The survey instrument was designed to achieve the objectives established in Chapter 1. Firstly, the questionnaire captures the main constructs in assessing the customer preference and perception, and the satisfaction in health care in various dimensions. The instrument assesses the ranks for motivation factor and health care package. It also measures a tourist’s travel characteristics and demographic profiles. The overall customer perception and satisfaction are tested with demographic background to identify the influencing attributes. Lastly, the questionnaire was used to test the effectiveness of retention technique and relation with overall satisfaction and perception.
3.1.3 Research Instrument

Customer survey in the ayurvedic health resort is the prime research instrument employed in this study. This primary data collection was supplemented by a spate source of secondary data. The secondary data pertaining to the study was gathered from books, journals, magazines, newspapers, articles, e-journals, web pages, various government department releases, brochures, pamphlets, conference, and symposium extracts.

In this study, the method used of collecting and gathering data from a part of population was by using a structured questionnaire. The questionnaire typically contains items each of which elicits a different bit of information. When a questionnaire that measures attitudes generally it must be constructed as an attitude scale and must use a large number of items in order to obtain a reliable assessment of an individual's attitude.

In this study, the measurement scales of questionnaire for all reflective items (perception and satisfaction) were based on a Five point Likert scale which is the common type of attribute measurement. Each of these items related to specific hypothesis. The scale ranged from very low to very high. The scale developed for customer perception varies from strongly disagree to strongly agree. This was used to indicate the degree of respondent’s agreement or disagreement with each item expressed in the form of the statements. This attitude measurement (Five point Likert) questionnaire was also used to evaluate the customer satisfaction ranging from highly satisfied to highly dissatisfied. Ranking questions were also administered to determine
the preference for selection of resort, selection of healthcare package, and selection of health care treatment.

The subsequent portion in the questionnaire is to assess the various customer retention tools practised in the health resorts. Finally, it gathers information on the socio economic profile of the health tourists.

In order to know the service rendered by the service providers, several rounds of discussions were held with resort managers, physicians and therapists to gain detailed informations on the topic studied.

3.1.4 Validity of Questionnaire*

In order to validate the designed instrument and remove any ambiguity among customers, the researcher conducted a pilot survey. For the purpose of pilot study, 50 respondents were consulted which included health tourists and service providers. The instrument designed was tested for reliability and validity before undertaking any full scale market survey. After conducting the pilot survey necessary, modifications were made in the questionnaire to conduct the present study. The reliability test revealed more than 70 per cent consistency in the responses on various indicators of health tourism helped to develop an unbiased scale. * Validity is the extent to which a scale or set of measures accurately represents the concept of interest (Hair et al., 2006).

3.1.5 Statistical Measurement

Quantitative data may be analysed using a variety of statistical packages such as Statistical Package for Social Sciences (SPSS). In this research, SPSS we used to
present the assumption, principles, and techniques necessary to gain insight into data. Several tests are used in this research includes. It Chi-square, weighted average, Garret ranking, ANOVA, factor analysis, discriminant function analysis, correlation path analysis, and regression analysis.

3.1.6 Target Population and Sample Size

The target population being all the health tourists who stayed and experienced the ayurvedic health care packages in the selected ayurvedic resorts of Kerala.

The sample size was set at 520 and 65 questionnaires each was distributed to the eight ayurvedic resorts under the study. Out of the 385 questionnaires, 25 questionnaires were incomplete and improperly done. Hence, 360 questionnaires were taken for the analysis of the present study. The method of the selection is convenience sampling, which was preferred for the study not because of saving cost and time but because of the convenience of health tourism operators and the guests. Another important hindrance the researcher faced was, the study was conducted during August 2008 to May 2009 when there was heavy economic slowdown and the Mumbai terror attack (November 2008).

3.1.7 Survey Procedure

A set of finalised questionnaire was distributed to all the survey locations (Eight ayurvedic resorts). The survey was conducted on site by the staff in the reception of the resort or ayurvedic consultant or with the therapist depending on the convenience of the service providers and the health tourists.
3.2 DATA ANALYSIS

Descriptive statistics determined the mean and standard deviation scores on attribute perception, attribute satisfaction, factor analysis, discriminant function, chi-square test, ANOVA for relationship testing, and Garret rating for ranking the various attributes. Discriminant analysis is used to identify the characteristics that can accurately discriminate between two attributes. Correlation analysis to identify patterns of variation common to a dependent variables and an independent variable.

3.2.1 Chi-Square Test

The Chi-square test is one of the simplest and most widely used non-parametric tests in statistical work. The symbol ($\chi$) is the Greek letter Chi. This test (S.L. Gupta, 2003) is used as a test of goodness of fit. Chi-square can be used in normal distribution as a test of independence. Test is done whether two or more attributes are associated (G.C. Beri, 2008). This test gives the magnitude of discrepancy between theory and observation. It is defined as

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where $O_i$ is observed frequency, $E_i$ is expected frequency.

In general expected frequency, is given by

$$E_i = \frac{RT \times CT}{N}$$

Here, $RT = \text{Row Total containing the cell.}$

$CT = \text{Column Total containing the cell.}$
\[ N = \text{Total Number of observations.} \]

**Degree of Freedom (V)**

The chi-square test is based on degree of freedom. This is obtained as

1. In case of one dimension (row or column) there are \((K-1)\) degrees of freedom, where \(K\) is number of categories of observed frequencies.

2. In case of contingency table
   \[ v = (r-1) (c-1) \]
   Here, \(r\) = is total number of rows
   \(c\) = is total number of columns

**Characteristics of chi square test**

1. This test is based on frequencies or events as against the ‘z’ and the ‘t’ test based on parameters like the mean and standard deviation.

2. This is applied for drawing inference only.

3. This possesses additive properties so that when \( (X_1^2) \) and \( (X_2^2) \) are independent and have a chi-square distribution with \( n_1 \) and \( n_2 \) degrees of freedom, it will also be distributed as a chi-square distribution with \( n_1 + n_2 \) degree of freedom.

4. It is a general purpose test and is very useful in research work.

Chi square analysis can be used when the data satisfy four conditions:
1. There must be two observed sets of data or one observed set of data and one expected set of data.

2. The two sets of data must be based on the same sample size.

3. Each cell in the data contains an observed or expected count of five or larger.

4. The different cells in a row or column must represent categorical variables.

3.2.2 ANOVA

One independent variable experiment is called one-way ANOVA. ANOVA stands for Analysis of Variance, the generic name given to a set of techniques for studying the cause and effect of one or more factors on a single dependent variable. Analysis of Variance technique is used when the independent variables are of nominal scale (categorical) and the dependent variable is metric (continuous), or at least interval scaled (Nargundkar, 2003). The ANOVA technique focuses on the behaviour of the variance within a set of data. In measures of dispersion, the variance of a variable is equal to the average squared deviation from the mean of the variable. The logic of the ANOVA technique says that if we calculate the variance between the groups and compare it with the variance within the groups. It makes a rational determination as to whether the means are significantly different (Hair et al., 2006).

It is a useful technique concerning researches in the fields of economics, education, business, and in other disciplines. This technique is used when multiple samples cases are involved. The significance of the difference between the means of
two samples can be found out through ‘Z’ test or t test, but the problem arises when it happens to examine the significance of difference amongst more than two sample means at the same time.

**Basic principles of analysis of variance (ANOVA)**

The basic principle is to test for difference among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples. In terms of variation within the given population, it is assumed that the values of \((X_{ij})\) differs from the mean of this population only because of random effects, i.e., there are influences \((X_{ij})\) which are unexplainable, whereas in examining differences between populations. It is assumed that the difference between the mean of the \(j^{th}\) population and the grand mean is attributable to what is called a 'specific factor' or what is technically described as treatment effect. Thus, while using ANOVA. It is assumed that each of the samples is drawn from a normal population and that each of these populations has the same variance. It is also assumed that all factors other than the one or more being tested are effectively controlled.

**The ‘F’ test**

*(Variance Ratio Test)*

The ‘F’-test is named in honour of the great statistician R.A.Fisher. It is used to find out whether the two independent estimates of population variance differ significantly. For carrying out the test of significance, we calculate the ratio ‘F’, ‘F’ is defined as:
\[ F = \frac{S_1^2}{S_2^2} \]

where
\[ S_1^2 = \frac{(X_1 - \bar{X}_1)^2}{n_1 - 1} \]
\[ S_2^2 = \frac{\sum (X_2 - \bar{X}_1)^2}{n_2 - 1} \]

note that \( S_1^2 \) is always greater than \( S_2^2 \) i.e., \( S_2^2 > S_2^2 \)

Therefore,
\[ F = \frac{\text{Larger estimate of variance}}{\text{Smaller estimate of variance}} \]
\[ V_1 = \text{degrees of freedom for sample having larger variance}. \]
\[ V_2 = \text{degrees of freedom for sample having smaller variance}. \]

The calculated value of ‘F’ is compared with the table value (given at 5 per cent or 1 per cent level of significance). If the calculated value of ‘F’ is larger than the table value of ‘F’, the null hypothesis is rejected otherwise it is accepted.

### 3.2.3 Discriminant Function Analysis

Discriminant analysis is a multivariate technique used for predicting group membership on the basis of two or more independent variables. Discriminant analysis is a technique for analysing marketing research data when the criterion or dependent variable is categorical and the predictor or independent variables are intervals. The linear combination of independent variables developed by discriminant analysis will best discriminate between the categories of the dependent variable (Hair et al., 2007). This
linear combination is called the discriminant function. It can be represented in the following way (Gupta, 2003).

\[ Z_i = b_1 X_{1i} + b_2 X_{2i} \ldots + b_n X_{ni} \]

- \( Z_i \) = \( i^{th} \) individuals discriminant score
- \( b_n \) = Discriminant coefficient for the \( n^{th} \) variable
- \( X_{ni} \) = Individuals value on the \( n^{th} \) independent variable.

Discriminant weights (\( b_n \)) or discriminant function coefficients are estimates of the discriminatory power of a particular independent variable.

The size of the coefficients associated with a particular independent variable is determined by the variance structure of the variables in the equation. Independent variables with large discriminatory power will have large weights and those with little discriminatory power will have small weights.

3.2.4 Factor Analysis

It is one of the more popular “analysis of interdependence” techniques. In studies of interdependence, all the variables are on an equal footing, and the analysis is concerned with the whole set of relationships among the variables that characterise the objects.

Unlike regression or discriminant analysis, factor analysis would focus on the whole set of interrelationships displayed by the variables. At the conceptual level, a factor is a qualitative dimension of the data that attempts to depict the “way in which
entities differ, much as the length of an object of the flavour of a product defines a qualitative dimension on which objects may or may not differ.

Factor analysis is a procedure that takes a large number of variables or objects and searches out factors in common which account for their inter-correlation.

**Application of Factor Analysis in Marketing Research**

There are a number of applications in marketing research. These include data reduction, structure identification, and scaling.

**Data reduction** : Factor analysis is used for reducing a mass of data to a manageable level. For example, the marketing researcher who has collected data on 40 attributes of a brand. This analysis and understanding of this data may be aided by reducing 30 attributes and having only 10 attributes. Factor analysis has been employed to “purify” original sets of scale items by isolating those items that do not reflect a common core and also to name the dimensions captured by measure. It is also used in the study of lifestyles and psychographic research problems in which it is used to develop consumer profiles that reflect people’s attitudes, activities, interests, opinions, perceptions, and preferences so as to better predict their consumption and purchase behaviour.

**Structure Identification** : Factor analysis may be used to understand the basic structure underlying a set of measures.

**Scaling** : A marketing researcher may develop a scale on a research topic. The elements in the factors are called unrotated factor loadings. The loadings measure
which variables are involved in which factor pattern, to what degree and in what
direction. They can be interpreted as correlation coefficients. The square of the loading
equals the proportion of the variation that a variable has in common with an unrotated
factor.

Another way to conceptualise this relationship is to remember that a loading is a
correlation coefficient between a variable and a factor. In essence, when we square a
loading we are calculating a coefficient of determination, between a variable and a
factor. Thus, the squared loading represents the amount of shared variation between a
variable and a factor.

The $L^2$ measures are called communalities. Communality is the proportion of a
variable’s total variation that is involved in the factors. Mathematically $L^2$ equals the sum
of the squared loading of a variable on all factors. To get the percentage of total
variance in the data explained by the four factors. We simply calculate $H$ when

$$H = \frac{\text{Sum of all } L^2\text{s}}{\text{Number of variables}} \times 100$$

This value is called the common variance explained by the factors (S.L.Gupta, 2003).

### 3.2.5 Garret Ranking

This technique was used to rank the motivating factors to select the health
services of this resort by the respondents under the study. In this method, the
respondents were asked to give ranks according to the magnitude of the factors. The
order of merit given by the respondents were converted into ranks by using the formula:
Percentage position = 100 \sum \left( \frac{R_{ij} - 0.5}{N_j} \right)

Where, \( R_{ij} \) = Rank given for 15\textsuperscript{th} factor by \( j \)\textsuperscript{th} individual

\( N_j \) = Number of factors ranked by \( j \)\textsuperscript{th} individual

The percentage position of each rank thus obtained is converted into scores by referring to the table given by Henry Garrett. Then for each factor the scores of individual respondents are added together and divided by the total number of respondents for whom scores were added. These mean satisfaction scores for all the factors are arranged in the descending order, ranks are given and most important problems are identified (Garrett, 2004).

### 3.2.6 Correlation and Path Analysis

Correlation analysis, like tabulation, attempts to identify patterns of variation comes to a dependent variable and an independent variable. Correlation analysis uses to examine the relationship between the variables. This results in an objectively arrived at correlation coefficient, which indicates low strongly the few variables share a common pattern of change and whether the pattern is positive or negative. The reasons of the relationship between two variables is the correlation co-efficient \( r \). If there is perfect positive correlation \( r = + 1.00 \); perfect negative correlation it is indicated by \( r = -1.00 \); no relationship is shown by \( r = 0.00 \). The equation used to calculate \( r \) is

\[
\begin{align*}
    r &= \frac{\sum [(Y_i - \bar{Y})(X_i - \bar{X})]}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (X_i - \bar{X})^2}} \\
    \text{where } Y_i &= \text{Value of dependent variables obtained from respondent } i. \\
    \bar{Y} &= \text{Average value of the dependent variables.}
\end{align*}
\]
\[ X_i = \text{Value of the independent variables obtained from respondents} \]

\[ \bar{X} = \text{Average value of the independent variables.} \]

(Boyd et al., 2004).

**Correlation Analysis**

In order to study the relationship between the independent variables and their influence on the dependent variable. The interconnection matrix is used when independent variables are in the X axis and dependent variables is on the Y axis. With the help of this matrix the relationship is explained (Boyd et al., 2004).

**Path Coefficient Analysis**

The path coefficient analysis showed the direct contribution of the explanatory variables as well its indirect contribution via other explanatory variables to the dependent variable. This relationship is well explained by matrix. The diagonal element in the matrix shows the direct effect of explanatory variables and the non-diagonal elements in the row shows the indirect effects of the explanatory variables through other variables and the sum of the distinct effect and inclined. Effects in the row accounts to the correlation coefficient of the explanatory variable with dependent variable.

**3.2.7 Regression Analysis**

A regression analysis is used to investigate the overall strength of the association or relationship in detail. A bivariate regression analysis is a statistical technique that users information about the relationship between an independent or predictor variable
and a dependent or criterion variable and combines it with the algebraic formula for a straight line to make predictions.

A fundamental basis of regression analysis is the assumption of a straight line relationship between the independent and dependent variables. The general formula for a straight line is (Hair et al., 2006).

\[ Y = a + bX + c_i \]

- \( Y \) = the dependent variable
- \( a \) = the intercept (point when the straight line intersects the \( Y \) and \( X \) when \( X = 0 \))
- \( b \) = the slope (the changes in \( Y \) for every vast changes in \( X \))
- \( X \) = the independent variable used to predict \( Y \).
- \( c_i \) = the error of the prediction.

In applying regression analysis, we examine the relationship between the independent variable \( X \) and the dependent variable \( Y \). After computing the values of \( a \) and \( b \), statistical significance test has to be done. The calculated \( a \) (intercept) and \( b \) (slope) are sample estimate of the true population parameters \( \alpha \) and \( \beta \). The ‘t’-test is used to determine whether the computed intercept and slope are significantly different from zero. ‘a’ is referred to as a constant and the ‘b’ is associated with each independent variable. Bivariate regression analysis is used when there is one independent variable and one dependent variable. Whereas multiple regression analysis is used when there is an several independent variables and one dependent variable. The regression analysis shows R square figures, which show the percentage of variation in one variable that is accounted for by another variable. The standard error
of the estimate is a measure of the accuracy of the prediction of the regression equation from the ANOVA table the ‘F’ ratio is the result of comparing the account of explained variance to the unexplained variance. The co-efficient table shows the regression co-efficient which is an indicator of the importance of independent variable in predicting a dependent variable. Once the statistical significance of the regression co-efficients is determined the relationship with the variables is identified.

3.3 CONCLUSION

The present study uses Garret Rating, weighted average and factor analysis as non-hypothesis tests to measure the preferences and group the similar clusters of perception and satisfaction on health tourism. A detailed analysis of socio-demographic factors with perception, satisfaction, and retention by using various tools such as chi-square, ANOVA, discriminant function, correlation, path analysis, and regression have been done to test various hypothesis evolved after the review of literature.