Chapter - 3

METHODOLOGY AND DATA BASE
Various analytical techniques, measures and tools were used to find out the answers for the questions raised in the form of objectives of this study. In this chapter, details of methodology followed are discussed under the following headings.

I. Study Area

Chikmagalur district was purposively selected among the four districts of major coffee growing districts of Karnataka state viz., Chikmagalur, Coorg, Hassan and Mysore, as chikmagalur district has the highest area under coffee cultivation in Chikmagalur district in Karnataka.

Chikmagalur and Mudigere taluks were again purposively selected among seven taluks in one district after taking into consideration the money and time limitation of the researcher. Moreover, taluks taken up for study have a sufficient number of coffee planters.

The present study was conducted in the Chikmagalur and Mudigere taluks of Chikmagalur district because in these taluks coffee is the main income generating source for small, marginal and large farmers and they mainly depend on coffee enterprise for their house hold income.

The random sampling procedure was adopted to select the sample respondent. A total of four villages was selected randomly out of 62 villages in Chikmagalur taluk and 78 villages in Mudigere taluk. From each of these eight villages, 30 planters were randomly chosen for the study to make up the predetermined 240 sample respondents. From each of these lists, four villages were selected randomly in each taluks.

From the selected eight villages the sample planters were selected randomly. The total size of the sample selected for the study was 240 respondents.
The planters were then classified into three groups, based on the size of the area under coffee cultivation viz., Small (0 to 5 acres), medium (5 to 10 acres) under coffee and large (10 and above acres) planters.

Multiple regression analysis was carried out to examine the factors associated and their relative contribution on coffee yield on various size groups of planters in Chikmagalur district in Karnataka during 2005-06.

The data relating to the output of coffee with the various levels of inputs in different size groups of planters are collected from the selected coffee producing area through a pre-testing questionnaire by personal enquiry during 2005-06.

II. Nature and Sources of Data

In order to evaluate the objectives of the study, data was collected from both primary and secondary sources.

a. Primary data

The data needed for the study were collected from the cultivators by personal interview method using, pretested questionnaires. The data records of the expenditure and income from coffee cultivation, were not maintained. Hence it was collected based on the memory of the respondent. At the time of interview, personal bias of the cultivators (planters) was minimized by convincing them about the genuine unity of the purpose for which the data was being collected. The data collected confined to fulfilling the objectives of the study from the selected cultivators. Data was based on the entire operations in establishing and maintaining the coffee plantations and the consequent costs and returns.

b. Secondary Data

The secondary data on area, production and productivity of coffee in different states of India and different district of Karnataka and different taluk of Chikmagalur district and only trend of production of coffee in major producing
countries were collected for a period of 10 years from 1995 to 2005 from different sources, namely District Coffee Board, District Statistics Office and UPASI (United Planters Association of South India), CCRI (Central Coffee Research Institute) etc.

The questionnaires used for data collection were designed based on the objectives of the study. The data so collected relate to:

- General characteristics such as name, education, qualification, occupation, land holding and its type, total area of the coffee plantations, bearing area, variety cultivated and name of the village and the taluk.
- Investment details for the establishment of coffee plantation, maintenance of coffee plantation, cost incurred in cultivation of coffee and intercrops, output obtained, quality of coffee produced and details on returns of the coffee and other intercrops, investment relating to irrigation equipments, farm equipment and cost structure.

III. Tabular Analysis

The technique of tabular analysis was employed for determining the investment pattern in the coffee plantations. This analysis was used to estimate the cost of the establishment of coffee plantation during non-bearing period, maintenance cost of coffee plantation during bearing period, cost structure, returns and profits. The tabulations were done separately for small, medium and large plantations. The parameters were worked out for per acre basis. The table was made separately for operation wise labour requirements, yield and returns.

IV. Establishment Cost

The money invested in the coffee plantations were indicated by the cost of establishment in the plantation. The establishment period of a coffee plantation is about three years in arabica coffee and from 6th year in robusta coffee. The costs
incurred in establishing a coffee plantation during the first three (Arabica) and six (Robusta) years were considered at prevailing wage rates and input prices. The establishment period up to three years and 6 years were taken as zero years. From the 4th and 6th year arabica and robusta the returns started considering the maintenance during the bearing period. The economic life of the coffee plantation is estimated to be 50-60 years. But for the purpose of computing the cost and returns, the economic life of the coffee was assumed at 50 years. That is three years and 6 years of establishment (zero years) and 47 years or 44 years of maintenance. The establishment cost was classified into material cost and labour cost.

Material cost included the cost incurred in zero year such as jungle clearing, seedling cost, line marking materials, fertilizers, PPC, shade plants, stakes, fuel, fencing materials etc.

Labour cost included the cost incurred in zero year such as jungle clearing, land preparation, line marking, staking, opening and filling pits, planting, fertilizer application, irrigation, shade regulation, topping, scuffling, mulching, hutting, trenching, harvesting, weeding and other miscellaneous items.

V. Maintenance Cost

The expenditure on cultivation included the cost of inputs and labour during bearing (yielding) period. The input used in the plantation was fertilizers, plant protection chemicals, agricultural lime, fuel and other miscellaneous/unforeseen expenses.

The actual expenditure incurred on labour for operation in coffee plantation during bearing period; fertilizer application, PPC application, irrigation, pruning, bush management, shade regulation. Manual weeding, liming, harvesting and processing etc., were used in computing the labour cost. In case where operations
were given on contract, the cost of contract was taken as maintenance cost for those cropping systems.

The labour wages for male and female was during the study period in some of the big estates of Chikmagalur districts where as labour wages varied, depending on the size of holding and location in some other estates. In this study labour wages accounted to determine the cost of cultivation was on an average at Rs.71/- per mandays, because the Indian Planters Association fixed the wage Rate at Rs.71/- per mandays.

Apart from the wages, planters in some of the estates had provided monthly salary to the appointed managers or mestrie, were also included in cost of cultivation of coffee plantation. Apart from these sorting and pulping costs were included in the harvesting and processing cost. However for the purpose of arriving at cost and returns in variety wise cropping systems, the processing cost of both parchment and cherry coffee preparations were average and included in the harvesting and processing charge component.

VI. Sampling Error

When we draw different samples from the population, it is possible that there will be some difference in their means, standard deviations etc. Statistical characters of population such as mean, standard deviation etc., are called parameters. The same characteristic features of samples which try to estimate the parameters of the population are known as estimates. The formula used to estimate these parameters are known as "Statistic". The differences between the values of the parameters and estimates are known as sampling errors. These errors relate solely to sampling fluctuations and not to human errors. The latter are known as non-sampling errors. When the sampling error is least, the sample is more representative of the universe. Larger the size of the sample lesser will be the sampling error.
VII. Standard Error

The standard error provides a measure to calculate the range within which sample statistic might deviate from the population parameter due to sampling error.

VIII. t-Distribution

Student's t-distribution

This distribution applicable to small samples was developed by W.S. Gossett who was an employee in a brewery. His employer did not permit him to get anything published in his name so he used a pen-name student and get his work published. This was done in the beginning of the present century. The distribution developed by Gossett, written under his pen name student has come to be known as student's t-distribution.

This distribution should be used when the sample size is less than 30. There is no mathematical line between large and small samples, but conventionally a sample of less than 30 items is considered small.

The student's t distribution is

\[ t = \frac{\overline{X} - \mu}{S / \sqrt{n}} \]

where,

S or the sample standard deviation is equal to

\[ S = \sqrt{\frac{\sum (X - \overline{X})^2}{n - 1}} \]

If sample S.D. is given without using n-1 as denominator then

\[ t = \frac{\overline{X} - \mu}{S / \sqrt{n-1}} \]
Once the value of 't' has been calculated it is compared with its table value. Here we do not use the normal distribution which is used in large samples. The table values of t have been calculated for various degrees of freedom at different confidence intervals.

IX. Variance Ratio Test or F-Test

The variance ratio test has a number of uses in statistical analysis. Here we will discuss only one important property of F-test i.e., to determine whether the two independent estimates of population variance significantly differ between themselves or whether they establish the fact that both the samples have come from the same universe and have a common variances. For this purpose the ratio of variance given by the two sample is obtained. This ratio is called F-ratio. Thus,

\[ F = \frac{S_1^2}{S_2^2} \]

where, \( S_1^2 = \frac{\sum (X - \bar{X})^2}{n_1 - 1} \) and \( S_2^2 = \frac{\sum (X_2 - \bar{X}_2)^2}{n_2 - 1} \)

In calculating F the numerator is the greater variance and the denominator the smaller variance. In other words,

\[ F = \frac{\text{Greater variance}}{\text{Smaller variance}} \]

The calculated value of F is compared with the critical value of F for the given level of significance and given degrees of freedom. If the calculated value of F > the critical value the difference is significant otherwise. It is insignificant and could have arisen due to fluctuations of sampling.

The degrees of freedom are indicated by \( V_1 \) and \( V_2 \) and are \( (n_1 - 1) \) and \( (n_2 - 1) \) respectively. The critical value of F of various degrees of freedom \( V_1 \) and \( V_2 \) are given by Snedecor for various levels of significance.
X. Technique of Analyzing Variance

The technique of analyzing the variance in case of two variables is similar. In both cases a comparison is made between the variance of sample means with the Residual variance. However in case of a single variable the total variance is divided in two parts only viz., variance between the samples and variance within the samples. The latter variance is the residual variance. In case of two variables the total variance is divided in three parts viz., i) Variance due to variables number one, ii) Variance due to variable number two and iii) Residual variance.

We shall discuss the analysis of variance separately for one variable and two variable or for one-way classification and two-way classification.

1. Variance Analysis in One-way Classification

In one-way classification we take into account only one variable - say the effect of different types of fertilizers on yield. Others factors like difference in soil fertility or the availability of irrigation facilities etc., are not considered. For one-way classification we may conduct the experiment through a number of sample studies. Thus, if 4 different fertilizers are being studied we may have 4 samples of say 10 fields each and conduct the experiment. We will note down the yield on each one of the field of various samples and then with the help of F-test try to find out if there is a significant difference in the mean yields given by different fertilizers.

We shall start with the Null hypothesis that is the mean yield of the 4 fertilizers is not different in the universe, or,

\[ H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \]

The alternative hypothesis will be

\[ H_0 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \]
With the above hypothesis we should calculate

i. The sum of the squares of variation between the samples.

ii. The sum of the squares of variation within the samples.

iii. The total sum of squares of variation.

iv. Calculate the F-Ratio.

v. Compare the F-Ratio so calculated with the critical value of F-Ratio.

vi. Draw inference whether the Null hypothesis is accepted or rejected.

**a. Variation between the Samples (SSC)**

The variation between the sample means can be either on account of difference in treatment (e.g. use of different type of fertilizers) or due the element of chance. The difference between the values of items in any single sample would be on account of chance because the same fertilizers has been used on all the fields included in that sample. The results of samples given different treatments (using different fertilizers) is given in columns and as such sum of squares of variation between the samples is called SSC (Sum of Squares in Column).

The degrees of freedom which is (C - 1) where C - No. of samples.

**b. Variation within the Samples (SSE)**

The variation within the samples would be on account of chances. Since the same treatment is given to all the items in the sample (the same fertilizer is being used on the sampled fields).

The difference in the values of various items in a sample which is due to chance is called an estimate of the error. It is denoted by SSE.

D.f. (N - C) as the number of columns would be equal to number of samples
c. **Total Sum of Squares**

This is the total of SSR and SSE. In other words, it is found by adding the sum of squares of deviations between the samples and the sum of squares of deviations within the samples.

The degrees of freedom which is $N - 1$, where $N$ stands for the total number of items in all the samples taken together. This is the total variance.

d. **Calculation of Variance Ratio**

Variance Ratio or $F$ is the ratio between greater variance and smaller variance. Generally variance between the samples is more than variance within the samples and $F$ Ratio is

$$F = \frac{\text{Variance between the Samples}}{\text{Variance within the Samples}} = \frac{\text{MSC}}{\text{MSE}}$$

The calculated $F$-ratio should be compared with the critical value of $F$ to draw inferences. One should be very careful in consulting the table containing the critical values of $F$. These values are given for various levels of significance on the basis of degrees of freedom for greater and smaller variance.

If the calculated value of $F$ is more than the critical value, the null hypothesis is rejected and the differences between the means is said to be significance. It means we accept the alternative hypothesis that the $\mu_1 \neq \mu_2 \neq \mu_3 \ldots$

All the points discussed above for analyzing variance if there is one way classification can be summarized in the following table which is sometimes referred to ANOVA table (Analysis of Variance table).
Chapter - 3

Methodology and Data Base

Analysis of variance table (One-way classification)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>F - Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between samples</td>
<td>SSC</td>
<td>C - 1</td>
<td>MSC = \frac{SSC}{C - 1}</td>
<td>MSC</td>
</tr>
<tr>
<td>Within samples</td>
<td>SSE</td>
<td>N - C</td>
<td>MSE = \frac{SSE}{N - C}</td>
<td>MSE</td>
</tr>
<tr>
<td>Total</td>
<td>SST</td>
<td>n - 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

where,

SSC = Sum of Squares between Samples
SSE = Sum of Squares within Samples
SST = Total Sum of Squares
MSC = Mean Square between Samples
MSE = Mean Square within Samples
F = Ratio of MSE to MSC

2. Variance Analysis in Two-way Classification

In a one-way classification we take into account the effect of only one variable. If there is a two-way classification the effect of two variables can be studied. The procedure of analysis in a two-way classification is total both the columns and rows. The effect of one factor is studied through the column wise figures and totals and of the other through the row wise figures and totals. The variances are calculated for both the columns and rows and they are compared with the residual variance or error. The table of variance analysis in a two-way classification takes the following form.
### Table: Analysis of Variance (ANOVA)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>Variance Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between samples</td>
<td>SSC</td>
<td>C (- 1)</td>
<td>MSC = (\frac{SSC}{C - 1})</td>
<td>MSC (\frac{MSC}{MSE})</td>
</tr>
<tr>
<td>Between Rows</td>
<td>SSR</td>
<td>r (- 1)</td>
<td>MSR = (\frac{SSR}{r - 1})</td>
<td>MSR (\frac{MSR}{MSE})</td>
</tr>
<tr>
<td>Residual error</td>
<td>SSE</td>
<td>(C (- 1))(r (- 1))</td>
<td>MSE = (\frac{SSE}{(C - 1)(r - 1)})</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SST</td>
<td>n (- 1)</td>
<td>(r (- 1))</td>
<td></td>
</tr>
</tbody>
</table>

where,

- \(SSC\) = Sum of Squares between Columns
- \(SSR\) = Sum of Squares within Rows
- \(SSE\) = Sum of Squares due to Errors
- \(SST\) = Total Sum of Squares

In a two way classification one should be careful in finding out the degrees of freedom. For column figures it is \((C - 1)\) for row total it is \((r - 1)\) and for the residual it is \((C - 1)(r - 1)\), while calculating F ratio. For columns and for rows the degrees of freedom for the numerator may not be the same.

### XI. Multiple Regression Analysis

This analysis is carried out to examine the factors associated and their relative contribution to the yield of coffee in various coffee producing districts as well as in the state. Per acre yield and various inputs used in coffee cultivation are collected from the various size groups of planters from the selected study area.

Using production function technique as the analytical tool, efforts are made to examine the relationship of coffee output with various cash inputs i.e., seedling, fertilizer, weeding, PPC, irrigation, the response of output to various inputs is estimated by functional analysis. The developed equation used in the study is:
\[ Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7 \text{ and } X_8) \]

where,
\[ Y = \text{Quantity of coffee produced per acre} \]
\[ X_1 = \text{Planted area} \]
\[ X_2 = \text{Bearing area} \]
\[ X_3 = \text{Coffee Production in Bags} \]
\[ X_4 = \text{Quantity of Labour} \]
\[ X_5 = \text{Quantity of Fertilizers} \]
\[ X_6 = \text{Quantity of PPC} \]
\[ X_7 = \text{Quantity of Irrigation} \]
\[ X_8 = \text{Transportation Cost} \]

XII. Estimates of Growth Rate

To find out the growth rates of area, production and productivity of coffee crop, the following form of trend equation is used in the state of India, different districts of Karnataka and different taluks of Chikmagalur and production of major coffee producing countries for the period 1995-96 to 2004-05.

\[ Y = AB_t V \quad (1) \]

where,
\[ Y = \text{Area or production or productivity in the year } 't' \]
\[ A \text{ and } B = \text{are parameters} \]
\[ A = \text{Intercept indicating the value of } 'Y' \text{ in the base period } 't' \]
\[ t = \text{time period starting from 1980-81} \]
\[ V = \text{Error term} \]
Percent of growth rate can be calculated as

\[ r = (\text{Antilog } B-1) \times 100 \text{ where,} \]

\[ r = \text{Average annual compound growth rate} \]

Equation (1) is converted into the logarithmic form in order to facilitate the use of linear equation. Taking logarithm on both sides of equation we get;

\[ \log A + t(\log B) + \log V \]

Or

\[ Y = a + bt + v \text{ ---------- (2)} \]

where,

\[ y = \log Y, b = \log B, a = \log A, v = \log V \]

The linear equation of above form is used for estimating annual compound growth rate of area, production and productivity of coffee crop grow in states of India. The value of ‘a’ and ‘b’ are estimated by using ordinary least squares method for the period 1995-96 to 2004-05.