CHAPTER-6
CHAPTER-6

BINARY EXPLANATORY VARIABLE APPROACH:
DISCRIMINANT ANALYSIS AND COMPARISON
OF SOCIO-ECONOMIC, DEMOGRAPHIC AND
SITUATIONAL CHARACTERISTICS OF 'POOR'
AND 'NON-POOR' HOUSEHOLDS

In the preceding chapter, we observed that in the same
cultural setting of the Rural Punjab, discernible
differences existed between the 'poor' and the 'non-poor'
cultivators. But on the basis of that limited exercise
it was not possible to generalize about the two groups in
terms of definite socio-economic characteristics that make
certain households highly prone to poverty and others
almost immune to it.

Accordingly, in this chapter, the search for
within-groups commonalty and between-groups differences
would be extended further to obtain a set of variables.....
sufficient to discriminate between the two groups. As a
first step, the two groups would be compared, taking one
characteristic at a time, through univariate analysis.
Subsequently, several characteristics would be considered
simultaneously and a search for a 'good discriminant
function' (capable of producing a very low degree of
misclassification for the known set of households, estimated
through step-wise discriminant analysis) will be made.

6.1 Comparison of Socio-Economic and Situational
Characteristics of 'Poor' and 'Non-Poor'
Households—a Univariate Approach

A large number of characteristics of the 'poor' and
'non-poor' farming units were available for making a comparative study of these groups. In one sense each observable characteristic merits attention and no list of dominant variables should be drawn a-priori. However, the vast theoretical literature on peasantry and scores of empirical studies in the area do provide justification for the inclusion of some specific characteristics regarding the size, structure, demographic mix, resource-endowment, allocative decisions and economic performance of units under study. Therefore, some fifty six variables have been included in the exploratory exercise for comparison of the chosen groups. For expository convenience, these variables have been grouped as follows:

I. Farm-size related variables;
II. Household size, demographic mix and location related variables;
III. Resource-quality related variables;
IV. Resource-use and allocative-efficiency related variables, and
V. Returns and Economic performance related variables.

A comparative study of the group means, for each of the variables involved, would be useful not only for taxonomic purpose but also for the purposes of gaining insight into processes of poverty generation and wealth accumulation in the population under study.
6.1(a) Testing for Differences Between Group Means—
'Dummy' (explanatory) Variable Approach

When it comes to testing for the statistical significance of difference between group means, with unknown population variance, we usually opt for the familiar student t-test or the z-test, as the case may be. Only in the case of comparison involving more than two groups, the accepted procedure requires an analysis of variance as a pre-requisite for further pair-wise comparison of the means.

Qualitative explanatory variables regression model provides us the same information besides being computationally more convenient and open to easy generalization. A brief description of the procedure follows:

Let X be some characteristic \( N(\mu, \sigma^2) \) for households belonging to either group, with \( \mu = \mu_1 \) for the 'poor' group and \( \mu = \mu_0 \) that for the 'non-poor'.

Then, formally, the situation can be described by

\[
X_i = \alpha + \beta D_i + u_i,
\]

where \( X_i \) is the magnitude of X for the ith household, and \( D_i \) is a 'dummy' (binary) variable with

\[
D_i = 1 \quad \text{if the unit is 'poor'},
\]

\[
= 0 \quad \text{otherwise};
\]

the term \( u_i \) is the disturbance variable which is assumed to satisfy all the basic assumptions of the classical normal linear regression model.
Then, $E(X_i|D_i = 0) = \alpha$
and $E(X_i|D_i = 1) = \alpha + \beta$

\[ \therefore \alpha = \mu_0, \text{ and } \alpha + \beta = \mu_1 \]
or $\beta = \mu_1 - \mu_0$ \hspace{1cm} \ldots (6.1.3)

which implies that the intercept of the population regression equation (6.1.1) measures the mean value of the 'non-poor' and the slope measures the difference between the mean value of a poor and a non-poor for any characteristic.

A test of the hypothesis that $\beta = 0$, would mean that the two group means do not differ. Now the coefficients $(\alpha, \beta)$ of the said regression equation can be estimated by the least squares method, obtaining the pair $(\hat{\alpha}, \hat{\beta})$, with $\hat{\alpha} = \bar{X}_0$ and $\hat{\beta} = \bar{X}_1 - \bar{X}_0$. Thus the least square estimate of the regression slope equals the difference between the sample means of the two groups. The $t$ or $z$-statistic can be appropriately utilized for testing the significance of difference between the two group means. In our case, since the degree of freedom is 250, application of $z$-statistic would be desirable.

6.1(b) Comparison of Socio-Economic and Situational Characteristics of 'Poor' and 'Non-Poor' Households—Results Based on the 'Dummy' Explanatory Variable Regression Exercise

The above described procedure was adopted for estimating fifty six regression equations, one each for the chosen variables to represent one or the other dimension of
### Table 6.1

**Comparison of the Farm Size-Related Variables: the Difference of Means Test for the 'Poor' and the 'Non-Poor' Households**

<table>
<thead>
<tr>
<th>Description of the Dependent Variables</th>
<th>Intercept ( \hat{\beta} )-Mean of the Control Group (i.e., Non-Poor)</th>
<th>Regression Coefficient ( \beta )-Mean of the other group (i.e., Poor) minus Mean of the Control Group</th>
<th>Z - Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of Crop Output (Rs.)</td>
<td>40277.56</td>
<td>-27732.80</td>
<td>3.86**</td>
</tr>
<tr>
<td>Value of Milk Output (Rs.)</td>
<td>5860.94</td>
<td>-2661.42</td>
<td>2.00*</td>
</tr>
<tr>
<td>Value of Farm Output (Rs.)</td>
<td>39945.74</td>
<td>-28344.20</td>
<td>3.84**</td>
</tr>
<tr>
<td>Area Owned (Acres)</td>
<td>11.09</td>
<td>-5.24</td>
<td>2.57*</td>
</tr>
<tr>
<td>Area Operated (Acres)</td>
<td>13.15</td>
<td>-6.71</td>
<td>2.98**</td>
</tr>
<tr>
<td>Number of Farm Workers</td>
<td>3.15</td>
<td>-0.64</td>
<td>2.08*</td>
</tr>
<tr>
<td>Value of Traditional Productive Assets (Rs.)</td>
<td>3145.37</td>
<td>-468.24</td>
<td>1.08</td>
</tr>
<tr>
<td>Value of Modern Productive Assets (Rs.)</td>
<td>19581.62</td>
<td>-13432.60</td>
<td>3.51**</td>
</tr>
<tr>
<td>Number of Milch Animals (Standard Units)</td>
<td>8.72</td>
<td>-3.08</td>
<td>2.47*</td>
</tr>
<tr>
<td>Total Human Labour (Man Days)</td>
<td>772.24</td>
<td>-314.51</td>
<td>3.08**</td>
</tr>
<tr>
<td>Value of Total Productive Assets (Rs.)</td>
<td>32772.83</td>
<td>-17925.80</td>
<td>3.41**</td>
</tr>
</tbody>
</table>

Source: Computed.  

* Significant at 5%  
** Significant at 1%
economic structure, size, organization, demographic composition and economic performance of the operating households.

The results are presented in Tables 6.1 through 6.5 under five group-headings (I, II, III, IV and V). Our first set pertains to farm size related variables. Size of an economic unit, being the most important characteristic having a direct bearing on the economic fate of a farm, has always figured in debates on poverty in rural areas. However, since there is no unique empirical measure of farm size itself, we employed for the purpose, eleven size-related variables. It is interesting to note that, but for the value of traditional productive assets, significant differences exist in the means for the two groups. Invariably, the 'non-poor' group showing a higher value for the mean, can be taken to mean that bigger is better—at least in the contemporary setting. Both, the stock as well as the flow variables display the same feature, thus strengthening Levinson's (1974) finding that stock and flow causes of misery cannot be dichotomised. Even if there are no economies of scale to be realized in agriculture, the bigger size holds a greater promise for keeping the households above subsistence level.

Table 6.2 throws further light on unequal living standards and possible causes of poverty. While there are
<table>
<thead>
<tr>
<th>Description of the Dependent Variables</th>
<th>Intercept</th>
<th>Regression Coefficient</th>
<th>Regression Coefficient β – Mean of the 'Poor' Group minus Mean of the 'Non-poor' Group</th>
<th>Z – Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Size* (Number of Heads)</td>
<td>7.54</td>
<td>0.11</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Family Size (Consumer Equivalent Units)</td>
<td>5.96</td>
<td>-0.05</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Percentage of Males in the Family</td>
<td>54.77</td>
<td>2.19</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>Worker-Consumer Ratio in the Family</td>
<td>0.40</td>
<td>-0.03</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Number of Workers in the Family</td>
<td>2.29</td>
<td>-0.11</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Cultivation Category of the Family</td>
<td>0.80</td>
<td>-0.10</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>Relative Economic Position of the Family</td>
<td>110.06</td>
<td>-58.02</td>
<td>2.74**</td>
<td></td>
</tr>
<tr>
<td>City-Gravity Variable</td>
<td>62.73</td>
<td>5.65</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Education level of the Head of the Household</td>
<td>2.74</td>
<td>-1.49</td>
<td>2.55*</td>
<td></td>
</tr>
<tr>
<td>Literacy Rate in the Unit</td>
<td>32.31</td>
<td>-2.17</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>Highest Education in the Family</td>
<td>7.03</td>
<td>-0.90</td>
<td>1.40</td>
<td></td>
</tr>
</tbody>
</table>

Source: Computed.  

* The terms Family, Household and Unit have been used interchangeably.  
* Significant at 5%  
** Significant at 1%
no significant differences in family size, worker-consumer ratio, number of workers in the family and percentage of males in the family, between two groups; the non-poor group possessed significantly higher stocks of wealth and flows of output/income. The implications are clear: the bigger is better, ceteris-paribus.

An insignificant difference in the means of urban influence variable is indeed surprising. For, if Lipton’s (1980) word on urban bias in development is accepted then one would expect the rural poor to be located somewhere in hinter-lands—far from the ‘trickle-down’ of urban development gains. On the other hand significantly higher ‘Relative Economic Position’ score for the non-poor group tends to suggest that greater control over the local factor markets does bring pecuniary gains as well. Though literacy rates as a whole do not differ across the groups but the poor units are headed by males with significantly lower levels of schooling than their counterparts in the non-poor group. While this difference can be taken to mean education as a poverty alleviating agent, the chain of causation could be the other way round.

So far as the resource quality related variables (Table 6.3) are concerned, it is only in the ownership of tractors and employment of hired human labour, where significant differences are observed. A relative scarcity
### Table 6.3
Comparison of Resource-Quality Related Variables

<table>
<thead>
<tr>
<th>Description of the</th>
<th>Intercept Mean of the Control Group</th>
<th>Regression Coefficient β - Difference of the Mean</th>
<th>Z - Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Literate Workers on the Farm</td>
<td>0.89</td>
<td>-0.26</td>
<td>1.93</td>
</tr>
<tr>
<td>Ownership of Draught-Power</td>
<td>0.82</td>
<td>-0.94</td>
<td>0.68</td>
</tr>
<tr>
<td>Possession of Own Sources of Irrigation</td>
<td>0.94</td>
<td>-0.03</td>
<td>0.78</td>
</tr>
<tr>
<td>Percentage of Area Irrigated</td>
<td>98.57</td>
<td>-0.97</td>
<td>0.78</td>
</tr>
<tr>
<td>Hired Human Labour (Equivalent Man-days)</td>
<td>475.18</td>
<td>-258.31</td>
<td>2.90**</td>
</tr>
<tr>
<td>Number of Plots per Holding</td>
<td>10.23</td>
<td>-2.09</td>
<td>0.90</td>
</tr>
<tr>
<td>Ownership of Tractor</td>
<td>0.29</td>
<td>-0.26</td>
<td>4.09**</td>
</tr>
</tbody>
</table>

Source: Computed

* Significant at 5%
** Significant at 1%
of labour power on the bigger farms or pure leisure-preference of the well-off might account for it. The two groups are hard to discriminate on account of percentage area irrigated, ownership of animal draught power and possession of wells and tube-wells.


However, at times, poverty is linked with lethargy, inadequate knowledge and inefficient patterns of resource utilization. Table 6.4 confronts these speculative assertions and popular notions with empirical realities. We find the poor employing more family labour per acre, devoting as much area to HIV crops (in percentage terms), adopting area specific cropping system and carrying out better diversified activities on the farm—findings contrary to the popular beliefs.

Low levels of productivity on the farms relative to those obtained in the area, as reflected by a significantly lower crop yield index, and non-availability of more remunerative urban employment happen to be the telling discriminators that tilt the scale against some families
### Table 6.4
Comparison of Resource Use and Allocative Efficiency Related Variables

<table>
<thead>
<tr>
<th>Description of the Dependent Variables</th>
<th>Intercept (Rs.)</th>
<th>Regression Coefficient (z-Values)</th>
<th>(\beta)-Difference of the Means</th>
<th>(z)-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Expenses</td>
<td>20945.98</td>
<td>-11153.80</td>
<td>3.10**</td>
<td></td>
</tr>
<tr>
<td>Percentage Area under MVY Crops</td>
<td>79.86</td>
<td>-3.01</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Intensity of Cultivation</td>
<td>191.51</td>
<td>-7.84</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>Use of Family Labour on the Farm (Man-days)</td>
<td>296.73</td>
<td>-55.88</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td>Family Labour as percentage of total Labour use</td>
<td>50.81</td>
<td>12.36</td>
<td>3.41**</td>
<td></td>
</tr>
<tr>
<td>Off-Farm Rural Employment (Number)</td>
<td>0.14</td>
<td>0.06</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Off-Farm Urban Employment (Number)</td>
<td>0.23</td>
<td>-0.16</td>
<td>2.15*</td>
<td></td>
</tr>
<tr>
<td>Total Off-Farm Employment (Number)</td>
<td>0.37</td>
<td>-0.10</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>Activity Diversification on the Farm</td>
<td>17.25</td>
<td>7.96</td>
<td>3.83**</td>
<td></td>
</tr>
<tr>
<td>Crop Concentration on the Farm (H - Index)</td>
<td>0.34</td>
<td>0.02</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Kharif Crop Area as % of total Crops Area</td>
<td>49.59</td>
<td>-1.63</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Use of Chemicals per Acre (Rs.)</td>
<td>366.01</td>
<td>-51.02</td>
<td>2.37*</td>
<td></td>
</tr>
<tr>
<td>Crop System Index</td>
<td>95.82</td>
<td>-2.76</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>Crop Yield Index</td>
<td>95.52</td>
<td>-18.90</td>
<td>4.63**</td>
<td></td>
</tr>
<tr>
<td>Use of Human Labour per Acre (Man-days)</td>
<td>67.96</td>
<td>5.24</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td>Paid-out Costs per Acre (Rs.)</td>
<td>1608.65</td>
<td>36.92</td>
<td>0.27</td>
<td></td>
</tr>
</tbody>
</table>

Source: Computed.

* Significant at 5%
** Significant at 1%

**Note:** The table lists various resource use and allocative efficiency-related variables, including operating expenses, area, intensity, and family labor usage, among others, along with regression coefficients and significance levels.
and leave them poor. In turn, lack of liquid resources, lesser use of micro plant nutrients, quality of productive base and, above all, the size of farms might explain lower levels of overall productivity on the farms. So, once again, size related variables come to the fore as powerful determinants of the current economic fate and consequently, good discriminators for deciding group membership.

The set of performance related variables (Table 6.5) unequivocally support low productivity doctrine of poverty. Significant differences in productivity are observed: low productivity going with the poor group and higher productivity with the non-poor. Hence small productive base, low levels of productivity and non-availability of better-paid jobs turn out to be strong correlates of poverty.¹³

But our results on growth of operations and expansion of ownership rights on land go against the dearly held immiserization and depeasantization hypotheses. We do not find any significant differences either in the renting-in or of net purchases of land between the poor and the non-poor groups. Myrdal’s oft-cited Biblical quote—“For to him who has more will be given, and he will have great plenty; but from him who has not, even the little he has will be taken away”—does not hold for our sample.¹⁴

In short, though the univariate analysis provides important clues to physiognomy of the poor and the non-poor
### Table 6.5

**Comparison of Returns and Economic-Performance Related Variables**

<table>
<thead>
<tr>
<th>Description of the Dependent Variables</th>
<th>Intercept μ-Mean of the Control Group</th>
<th>Regression Coefficient β-Difference of Means</th>
<th>Z-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income from the Sale of Labour (Rs.)</td>
<td>1253.81</td>
<td>-590.17</td>
<td>1.77</td>
</tr>
<tr>
<td>Income from Machinery Hiring-out (Rs.)</td>
<td>517.75</td>
<td>-517.75</td>
<td>1.63</td>
</tr>
<tr>
<td>Income from Land Rented-out (Rs.)</td>
<td>306.09</td>
<td>-175.18</td>
<td>0.72</td>
</tr>
<tr>
<td>Net Household Income (Rs.)</td>
<td>19683.14</td>
<td>-17549.20</td>
<td>4.60**</td>
</tr>
<tr>
<td>Current Debt (Rs.)</td>
<td>3238.76</td>
<td>-413.30</td>
<td>0.33</td>
</tr>
<tr>
<td>Percentage Area added through Renting-in Land</td>
<td>15.86</td>
<td>-6.67</td>
<td>1.13</td>
</tr>
<tr>
<td>Net Purchases of Land (Acres)</td>
<td>0.19</td>
<td>-0.78</td>
<td>1.70</td>
</tr>
<tr>
<td>Farm Output per Acre (Rs.)</td>
<td>3027.60</td>
<td>-885.45</td>
<td>4.74**</td>
</tr>
<tr>
<td>Farm Output per Man-day (Rs.)</td>
<td>47.59</td>
<td>-18.27</td>
<td>6.67**</td>
</tr>
<tr>
<td>Farm Output per Rupee spent on Operations (Rs.)</td>
<td>2.04</td>
<td>-0.52</td>
<td>3.70**</td>
</tr>
<tr>
<td>Total Off-Farm Income (Rs.)</td>
<td>2077.65</td>
<td>-1283.10</td>
<td>2.59**</td>
</tr>
</tbody>
</table>

*Source: Computed.*

* Significant at 5%
** Significant at 1%
cultivating households, yet it remains to be seen if the correlated variables involved, taken simultaneously, obtain a sufficient degree of discrimination between the two groups. We probe into this aspect with the use of multi-variate analysis technique in Section 6.2.

6.2 Comparison of 'Poor' and 'Non-Poor' Cultivating Households through Discriminant Function Analysis

While comparing the 'poor' and 'non-poor' groups on the basis of measurements on several characters: $X_1, X_2, \ldots, X_p$, we are interested in two issues:

1) How far, if at all, does the multivariate data confirm the prior categorization of the individual households into the two groups? and

2) Construction of a rule, based on the same set of measurements, which should enable us to assign a new observation to either of the two groups without any serious degree of misclassification.

The main motivation for undertaking this type of work came from the reported high degree of misidentification in specific beneficiary-oriented schemes covered under India's anti-poverty programmes. While it is very difficult to check an intentional leakage of funds allocated for such schemes, a definite objective criterion, constructed for the purpose of classification, can always minimize the unintentional leakages.
Use of a notional poverty line cannot be of universal help since no dependable records of rural households' income exist. It is, therefore, desirable to find a rule, based on directly identifiable objective characteristics of these households, which is capable of sufficiently discriminating between the poor and the non-poor besides producing a fairly correct classification.

To our knowledge, little work has been done in this direction. Accordingly, the present exercise would offer useful clues for identifying the poor and, hence, prove a good tool for any meaningful assault on poverty. Discriminant analysis ideally suits our purpose since it helps in answering both the questions of our interest. The main features of Fisher's linear discriminant function for the two-group case, relevant to our problem, are described below.

6.2(a) Discrimination and Classification—
Fisher's Discriminant Function

Fisher (1936) formally proposed the linear discriminant function as a solution to a practical problem of achieving optimal separation of two species of plants using a number of correlated variables. Ever since, the techniques of discriminant function, due to Fisher, the multivariate test statistic, $T^2$, due to Hotelling and the distance function, $d^2$, due to Mahalanobis have been established to be alike, and the technique of discriminant analysis has become popular.
for tackling the twin problems of discrimination and classification involving two (or more) groups.

Let \( X^1, X^2, \ldots, X^p \) be the \( p \) variables on which measurements are available for all the cases belonging to either group; \( N_g \) be the number of cases in the \( g \)th group, \( g = 1, 2 \) be the two groups. Let \( X^g_{ij} \) be the value of the \( j \)th variable of the \( i \)th observation from the \( g \)th group, where

\[
l = 1, 2, \ldots, N_g \]
\[
g = 1, 2 \]
\[
j = 1, 2, \ldots, p.\]

The sample mean vectors for the two groups are

\[
X^1 = \begin{bmatrix} X^1_{11} \\ X^1_{12} \\ \vdots \\ X^1_{1p} \end{bmatrix} \quad \text{and} \quad X^2 = \begin{bmatrix} X^2_{11} \\ X^2_{12} \\ \vdots \\ X^2_{1p} \end{bmatrix}
\]

... \(6.2.1\)

where \( X^g_{ij} = \frac{1}{N_g} \sum_{l=1}^{N_g} X^g_{ij} \).

The difference between the two sample mean vectors is:

\[
d = X^1 - X^2 = \begin{bmatrix} X^1_{11} - X^2_{11} \\ X^1_{12} - X^2_{12} \\ \vdots \\ X^1_{1p} - X^2_{1p} \end{bmatrix}
\]

... \(6.2.3\)
The estimated within samples variance-covariance matrices for the two groups, on \((N_1 - 1)\) and \((N_2 - 1)\) degrees of freedom, respectively, can be taken as

\[
\frac{1}{N_1 - 1} S(1) \quad \text{and} \quad \frac{1}{N_2 - 1} S(2)
\]

where \(S(1)\) and \(S(2)\) are the sum of product matrices whose elements are defined as follows:

\[
S_{ij}(1) = \sum_{l=1}^{N_1} (X_{1lj} - \bar{X}_{1i})(X_{1lj} - \bar{X}_{1j})
\]

and

\[
S_{ij}(2) = \sum_{l=1}^{N_2} (X_{2lj} - \bar{X}_{2i})(X_{2lj} - \bar{X}_{2j})
\]

\((i, j = 1, 2, \ldots, p)\).

Furthermore, the estimated pooled variance-covariance matrix \(S\), for the two groups with \(N_1 + N_2 - 2\) degrees of freedom is given by

\[
S = \frac{1}{N_1 + N_2 - 2} [S(1) + S(2)]
\]

Fisher's discriminant function between two populations is defined as that linear combination of variables \(X_1, X_2, \ldots, X_p\) for which the ratio of the square of the difference of its expectations in the two groups to its variance is maximized, i.e., it is that combination,

\[
Z = L'X, \quad \text{for which}
\]

\[
M = \frac{L'dd'L}{L'SL},
\]

is maximum, where \(L' = [\lambda_1, \lambda_2, \ldots, \lambda_p]\) is the vector
of discriminant coefficients. For maximization of $M$, the partial derivative of $M$ w.r.t. $L$ gives,

$$\frac{\partial M}{\partial L} = \frac{2[L'SL][dL'd] - 2[L'd'SL][SL]}{[L'SL]^2} \quad \ldots (6.2.7)$$

setting $\frac{\partial M}{\partial L} = 0$, we obtain

$$2[L'SL][dL'd] = 2[L'd'SL][SL] \quad \ldots (6.2.8)$$

i.e., $dL'd = MSL \quad \ldots (6.2.9)$

when $d$ is a non-zero vector, and $L'd \neq 0$. Pre-multiplying both sides by $M^{-1}$ gives

$$SL = M^{-1}dL'd. \quad \ldots (6.2.10)$$

Without loss of generality, we put $M^{-1}dL' = 1$ and obtain

$$L = S^{-1}d \quad \ldots (6.2.11)$$

The Mahalanobis distance statistic, $D^2$, is then defined as

$$D^2_p = d'S^{-1}d = d'S^{-1}d = d'L \quad \ldots (6.2.12)$$

6.2(b) An Alternative Presentation of the Discriminant Values

An adaptation of Fisher's (1936) procedure permits the derivation of a separate linear combination, called a 'classification function' for each group. These functions have the following form:

$$w_g = c_{g0} + c_{g1}X_1 + \ldots + c_{gp}X_p \quad \ldots (6.2.13)$$
where \( w_g \) is the score for group \( g \) and \( C_{gj} \)'s are the coefficients needed to be derived by the following computation (Klecka, 1980, p.43)

\[
C_{gj} = (n-2) \sum_{i=1}^{P} a_{ji} x_{ig}, \quad \ldots (6.2.14)
\]

where \( C_{gj} \) is the coefficient for variable \( j \) in the equation corresponding to group \( g \), and \( a_{ji} \) is an element from the inverse of the within-groups sum of cross-products matrix. The required constant term is defined as

\[
C_{g0} = \frac{-\left( \sum_{i=1}^{P} C_{gi} x_{ig} \right)}{2} \quad \ldots (6.2.15)
\]

These classification function coefficients need not be interpreted directly. However, following Bennett and Bowers (1976, pp.103-4), one can compute \( \lambda_j \)'s from \( C_{gj} \)'s.

We had to rely on this procedure because of its availability on the computer package.

6.2(c) **Step-wise Discriminant Analysis**

In both of the previous approaches, the discriminant functions may be created from the entire set of independent variables, irrespective of the discriminating power of each of the variables. An alternative procedure is to use a step-wise selection method, wherein independent variables are selected for entry into the analysis on the basis of their discriminating power.
The effects of dimension in discriminant analysis have been thoroughly considered by Van Ness and Simpson (1976) on the basis of which Srivastava and Carter (1983, p.243) observe that "use of fewer variables makes the discriminant function easier to work and allow better estimates of the parameters remaining in the model." Therefore, it is desirable to select a subset suited to a given research situation. Ideally, one would like to look at all possible subsets of variables and test their relative efficiency by examining either i) the reduction of the between group variances with respect to the within group variances, using an F-statistics; or ii) the decrease in the correct classification rate.

But this would be too expensive an exercise. So we adopted a two stage procedure: creating five sub-groups of variables at the first stage and performing a step-wise discriminant analysis separately for each combination at the second stage, using EMD computer package. The program uses a step-wise F-statistic criterion. At each step, until a threshold value for the F-statistic is reached, the variable that will produce the maximal F-value is added. Variables are then removed by deleting at each succeeding step the variable that will produce the minimal F-value.

Even, before a variable is considered for selection, it has to qualify the 'tolerance' test. The idea being to
preserve computational accuracy and search for unique information about groups (Klecka, 1980, pp.56-58). The tolerance for a variable not yet selected is one minus the squared multiple correlation between that variable and all variables already entered, when the correlations are based on the within groups correlation matrix. If the variable being tested is a linear (or nearly so) combination of one or more of the included variables, its tolerance will be zero (or near zero) and the variable will not even be considered for selection.

Besides this tolerance-test, the partial multivariate F-statistics, 'F-to-enter' and 'F-to-remove' test the change in discrimination due to a particular variable. While the former tests the additional discrimination introduced by the variable, the latter tests the significance of the decrease in discrimination, should that variable be removed from the list of variables already selected.

Habbema and Hermans (1977) have spelled out the disadvantages of this F-statistic criterion for more than two groups. But in our exercise only two groups are involved, so the procedure works very well. Hence, those variables which qualify the tolerance test and assure that the increased discrimination exceeds some chosen level, as per the partial F-statistic, can be subjected to an equivalent selection criterion: the variable selected is the one which after entering minimizes
i.e., the groups which are close together get separated on the inclusion of that variable.

6.2(d) Testing the Significance of the Discriminant Function

Having obtained a discriminant function we would, obviously, be interested in knowing if it can significantly distinguish between the two groups. A quick approximate method is to use the well-known test of significance for the difference between the means of two normal distributions (Bennett and Bowers 1976, pp. 99-100). But the normal test is only a rough guide since it takes no account of the number of variables involved.

A more exact test is given by Rao (1973, p. 480). It is the 'variance-ratio', based on $D^2_p$, which has an $F$-distribution with $v_1$ and $v_2$ degrees of freedom. The variance ratio is given by:

$$
P = \frac{N_1 + N_2 - p - 1}{p} \frac{N_1 N_2}{(N_1 + N_2)(N_1 + N_2 - 2)} \frac{D^2}{p} \quad \text{...(6.2.17)}$$

where $N_1$ = the number of cases in group one,

$N_2$ = the number of cases in group two,

$v_1 = p$ (the number of variables included)

$v_2 = N_1 + N_2 - (p+1)$.

However, when $v_2$ is large, the statistic $\frac{N_1 N_2 D^2}{N_1 + N_2 p}$ is asymptotically distributed chi-square with $v_1$ degrees of freedom, and the same should be used as the test-statistic.
The null hypothesis of no difference between the two groups, on the basis of multiple inter-related characteristics of cultivating households, can be tested at any pre-assigned acceptable level of significance.

6.2(e) Classification on the basis of Posteriori Probabilities

From equation (6.2.13) we can obtain the value of the 1st classification function evaluated at case 1 of group 2 as

\[ w_{21l} = c_{10} + \sum_{j=1}^{p} c_{1j} x_{1lj}, \]

\[ l = 1, 2, \ldots \ldots, N_2 \]

and vice-versa. Then, under the normality assumption, posteriori probability of case 1 in group 2 having come from group 1 is given by

\[ p_{211} = \frac{\exp(w_{211})}{\sum_{g=1}^{2} \exp(w_{21g})} \]

These probabilities sum to 1.0 over the two groups and classification on the largest of these values is also equivalent to using the smallest distance.26

The above described method of classification assumes that all the groups weigh equally in our scheme of things. However, in certain situations, we might have authentic information on probabilities of group membership—i.e., prior probabilities. So adjustment of posteriori probabilities, to account for prior knowledge of likely group membership, would reduce the chances of misclassification.
This adjustment is made to the simple classification functions by adding the natural logarithm of the prior probability, for that group, to the group's constant term.27

The above stated procedure has been used for discriminating between the 'poor' and the 'non-poor' groups and also for classifying known cases into the two groups. The results are discussed in the section that follows.

6.3 **Discriminant Analysis and Classification of the Cultivating Households into 'Poor' and 'Non-Poor' Groups.**

In the study in hand, we had fifty-five known cases of the poor and one hundred ninety seven cases of the non-poor group. We intend to examine that, how far, if at all, the multivariate data confirms the prior categorization of cultivating households into the two groups?

Furthermore, we are also interested in constructing a rule which would enable us to assign a new observation to either of the two groups, without our having to depend upon any information about the household income/expenditure. On the basis of our univariate exercise, we have selected the following variables for conducting this multi-variate exercise (i.e., Discriminant Analysis) which would help us in solving both the problems,

\[ X_1 = \text{Area Operated (acres)}; \]
\[ X_2 = \text{Number of Family Farm Workers (Male equivalent units)}; \]
\[ X_3 = \text{Traditional Productive Assets (Value in Rs.)}; \]
\[ X_4 = \text{Modern Productive Assets (Value in Rs.)}; \]
\[ X_5 = \text{Total Operating Expenses (Value in Rs.)}; \]
\[ X_6 = \text{Number of Milch Animals (Thousand rupee equivalents)}; \]
\[ X_7 = \text{Family Size (Number of heads)}; \]
\[ X_8 = \text{Percentage of Males in the Family}; \]
\[ X_9 = \text{Worker Consumer Ratio in the Household}; \]
\[ X_{10} = \text{Education of the Head of the Household (Number of years of formal education)}; \]
\[ X_{11} = \text{Relative Position of the Unit}; \]
\[ X_{12} = \text{Percentage of land area 'sucked-in'}; \]
\[ X_{13} = \text{Percentage Area under High Yielding Varieties (HYV Seeds)}; \]
\[ X_{14} = \text{Intensity of Cultivation}; \]
\[ X_{15} = \text{Family labour as percentage of total human labour used on the farm}; \]
\[ X_{16} = \text{Off-farm Rural Employment (Number of persons)}; \]
\[ X_{17} = \text{Off-farm Urban Employment (Number of persons)}; \]
\[ X_{18} = \text{Use of Chemicals per acre (Rs.)}; \]
\[ X_{19} = \text{Activity Diversification on the farm or Dairy Output as a percentage of the total farm output}; \]
\[ X_{20} = \text{Value of farm output per acre (Rs.)} \]

The rationale for using these variables to represent suitable characteristics of farming units has already been discussed in detail in Chapter 4 of the current study.
In all five alternative combinations, of different variables, were considered for the purposes of obtaining significant discrimination between the two groups. The results are shown in Table 6.6. Following Klecka (1980), prior probabilities of 0.19 and 0.81, for the 'poor' and 'non-poor' groups respectively, were used in obtaining the classification matrix through alternative combinations of variables.

The accuracy of the classification procedure has been tested by applying the classification rule to known cases packed in the sample. An idea of predictive accuracy of these functions could be had from 'proportional reduction in error statistic', \( \bar{\tau} \), defined as

\[
\bar{\tau} = \frac{N_c - \sum_{g=1}^{2} p_g N_g}{N - \sum_{g=1}^{2} p_g N_g} \quad \ldots \quad (6.3.1)
\]

where \( N_c \) = the number of cases correctly classified;

\( p_g \) = the probability of the gth group membership;

\( N_g \) = the number of cases in the gth group;

\( N = \sum_{g=1}^{2} N_g \).

The value of \( \bar{\tau} \) (Tau) lies between 0 and 1, the maximum value, 1, corresponding to no errors in prediction and that of 0, (the minimum), indicating that the classification is as good as any random procedure.
Table 6.6
Performance of Alternative Linear Discriminant Functions:
Distance between 'Poor' and 'Non-Poor' Households and
Predictive Accuracy

<table>
<thead>
<tr>
<th>Discriminant Function</th>
<th>Characteristics Included in the Discriminant Function</th>
<th>No. of Characteristics</th>
<th>Mahalanobis Distance ((\chi^2))</th>
<th>Chi-square Statistic (p)</th>
<th>Standardized Error Statistic (\tau)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.Fn - I</td>
<td>(X_1, X_2, X_3, X_5, X_6, X_7, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19}, X_4)</td>
<td>18</td>
<td>2.07191</td>
<td>89.0834**</td>
<td>1.0000</td>
</tr>
<tr>
<td>D.Fn - II</td>
<td>(X_1, X_2, X_3, X_5, X_6, X_7, X_9, X_{10}, X_{13}, X_{14}, X_{16}, X_{17}, X_{18}, X_4)</td>
<td>12</td>
<td>1.00506</td>
<td>43.2135**</td>
<td>0.9524</td>
</tr>
<tr>
<td>D.Fn - III</td>
<td>(X_1, X_2, X_6, X_9, X_{10}, X_{13}, X_{14}, X_{16}, X_{17}, X_{18}, X_4)</td>
<td>11</td>
<td>0.99297</td>
<td>42.6937**</td>
<td>0.9683</td>
</tr>
<tr>
<td>D.Fn - IV</td>
<td>(X_1, X_2, X_3, X_5, X_6, X_9, X_{10}, X_{14}, X_{15}, X_{16}, X_{17}, X_4)</td>
<td>10</td>
<td>0.69270</td>
<td>29.7859**</td>
<td>0.9048</td>
</tr>
<tr>
<td>D.Fn - V</td>
<td>(X_1, X_2, X_6, X_9, X_{10}, X_{14}, X_{16}, X_{17}, X_4)</td>
<td>9</td>
<td>0.68320</td>
<td>29.3792**</td>
<td>0.8571</td>
</tr>
</tbody>
</table>

Source: Computed.

Figures in brackets are degrees of freedom.
* Significant at 5%.
** Significant at 1%.
The results show that the discriminant function \( D_{F_n} - I \) produces a highest \( D^2 = 2.07191 \), significant at 1 per cent level, and the maximum \( \tau = 1 \). In contrast, the discriminant function \( D_{F_n} - V \) results in \( D^2 = 0.8630 \), again significant at 1 per cent level, and a lower value of \( \tau = 0.8571 \). Performance of the other discriminant functions varied between these two extremes.

On the face of it, the function \( D_{F_n} - I \), which involves eighteen independent variables is the best choice. But given our objective of finding an observable and operationally convenient set of discriminators, this function is the least acceptable one. For, though the selected variables cover all the aspects of a farm economy and collectively produce a zero degree of misclassification, but their use creates several computational problems. The quantum of information required here can be straight away used to obtain an estimate of net-household income itself. That defeats the very purpose of our exercise. In addition, the inclusion of size, productivity, input-use and households' characteristics in the same go may not be to the liking of all scholars.\(^{30}\) Exclusion of certain variables from the complete set, under the above stated constraints, saw us forming four other sets of variables resulting in other discriminant functions. However, in the process of elimination, some information must have been lost. Therefore, the variance ratio test with
on \( p-q \) and \( (N_1+N_2-p-1) \) degrees of freedom, has been conducted. It amounts to testing the hypothesis that the excluded variables, numbering \( p-q \), do not provide additional discrimination.

Results of test for additional information are presented in Table 6.7. A perusal of the results shows that exclusion of variable \( X_5 \) (Total operating expenses on the farm), from \( D.Fn -II \) to obtain \( D.Fn -III \) or that from \( D.Fn -IV \) to obtain \( D.Fn -V \), does not result in any significant loss of information. But in all other cases a reduction in the number of included variables is accompanied by a significant loss of discrimination between the two groups.

So the statistical implications of our decision are clear—dropping variables from the complete set does involve some loss of information. Consequently, a puritan interested in discrimination, per-se, would opt for \( D.Fn -I \). The discriminant function \( D.Fn -V \) would rank fifth in his scheme of things. However, a policy maker interested in identifying the poor on the basis of self-evident characteristics, with a minimal use of informational resource, should be happy working with \( D.Fn -V \). More so, since its use ensures eighty six per cent fewer chances of making errors, as compared with any arbitrary classification, in allocating...
<table>
<thead>
<tr>
<th>Comparison of 'Distance for the Linear Discriminant Functions</th>
<th>Change in Distance $\Delta D^2 = D^2_p - D^2_q$</th>
<th>The Variance Ratio: $F(p-q, N_1 + N_2 - p - 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.Fn - I $\neq$ D.Fn - II</td>
<td>$D^2_{18} - D^2_{12} = 1.06685$</td>
<td>$6.07556^{**}$ (6,233)</td>
</tr>
<tr>
<td>D.Fn - II $\neq$ D.Fn - III</td>
<td>$D^2_{12} - D^2_{11} = 0.01209$</td>
<td>$0.42449$ (1,239)</td>
</tr>
<tr>
<td>D.Fn - III $\neq$ D.Fn - IV</td>
<td>$D^2_{11} - D^2_{10} = 0.30021$</td>
<td>$11.07322^{**}$ (1,240)</td>
</tr>
<tr>
<td>D.Fn - IV $\neq$ D.Fn - V</td>
<td>$D^2_{10} - D^2_{9} = 0.00946$</td>
<td>$0.35087$ (1,241)</td>
</tr>
<tr>
<td>D.Fn - I $\neq$ D.Fn - III</td>
<td>$D^2_{18} - D^2_{11} = 1.07894$</td>
<td>$5.27599^{**}$ (7,233)</td>
</tr>
<tr>
<td>D.Fn - I $\neq$ D.Fn - IV</td>
<td>$D^2_{18} - D^2_{10} = 1.37915$</td>
<td>$6.17331^{**}$ (8,233)</td>
</tr>
<tr>
<td>D.Fn - I $\neq$ D.Fn - V</td>
<td>$D^2_{18} - D^2_{9} = 1.38861$</td>
<td>$5.53304^{**}$ (9,233)</td>
</tr>
<tr>
<td>D.Fn - II $\neq$ D.Fn - IV</td>
<td>$D^2_{12} - D^2_{10} = 0.31230$</td>
<td>$5.73558^{**}$ (2,239)</td>
</tr>
<tr>
<td>D.Fn - II $\neq$ D.Fn - V</td>
<td>$D^2_{12} - D^2_{9} = 0.32176$</td>
<td>$3.94528^{**}$ (5,239)</td>
</tr>
<tr>
<td>D.Fn - III $\neq$ D.Fn - V</td>
<td>$D^2_{11} - D^2_{9} = 0.30967$</td>
<td>$5.71959^{**}$ (2,240)</td>
</tr>
</tbody>
</table>

Source: Computed

** Significant at 1%.
A classification of the known cases, with D.Pn-V resulted in the following classification matrix:

<table>
<thead>
<tr>
<th>Known Group Membership</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
</tr>
<tr>
<td>Poor (55)</td>
<td>50</td>
</tr>
<tr>
<td>Non-Poor (197)</td>
<td>13</td>
</tr>
</tbody>
</table>

Taken as a whole, only 7.14% cases have been misclassified on the basis of D.Pn-V. If such a degree of error is not acceptable then D.Pn-III, which involves eleven variables in all, is worthy of consideration. This function correctly classifies 54 out of 55 poor and 194 non-poor from amongst 197 known cases i.e., an overall 1.59% misclassification is observed. However, this formulation includes variable $X_{49}$ i.e., activity diversification on the farm whose computation requires (i) value of agricultural produce, (ii) value of crops and by-products fed to animals, and (iii) value of dairy output. We can well imagine the informational costs and operational hazards of this prescription.

It seems that, from the operational point of view, the function D.Pn-V is the most logical choice under these conditions, even if some degree of misclassification is there. Hence, a collection of the variables $X_1$, $X_2$, $X_6$, $X_9$, $X_{10}$, $X_{14}$, $X_{16}$, $X_{17}$ and $X_{18}$ can be retained for
classifying cultivating households into 'poor' and 'non-poor' groups. The following functions emerge from our data:

For Group A (Non-Poor)

\[ 0.17938X_1 - 0.77485X_2 + 0.0057X_6 + 17.54016X_9 \\
-0.05994X_{10} + 0.32796X_{14} + 1.23175X_{16} + 0.65513X_{17} \\
-0.01151X_{18} - 33.17759. \]  

(6.3.3)

For Group B (Poor)

\[ 0.13662X_1 - 0.47212X_2 - 0.00383X_6 \\
+ 15.54300X_9 - 0.11023X_{10} + 0.31816X_{14} \\
+ 1.08969X_{16} - 0.27572X_{17} - 0.01301X_{18} \\
- 30.01846. \]  

(6.3.4)

From these two functions we get the required discriminant function as

\[ Y = 0.04276 \text{(Area Operated)} + 0.30273 \text{(Number of family farm workers)} + 0.00442 \text{(Number of Milch Cattle)} + 1.99716 \text{(Worker Consumer Ratio)} + 0.05029 \text{(Education Head)} + 0.00980 \text{(Intensity of Cultivation)} + 0.14206 \text{(Off-farm rural employment)} + 0.93085 \text{(Off-farm urban employment)} + 0.00150 \text{(Use of chemicals per acre)}. \]  

(6.3.5)

At this stage some comments about the relative contribution of individual variables towards overall discrimination are in order.
In the literature, two procedures are available for the purpose. The use of 'F-to-remove' statistic can be made to obtain the rank order of the unique discriminating power carried by each of the selected variables (Klecka, 1980, pp. 57-58). The variable with the largest F-to-remove is accepted as making the largest contribution to overall discrimination above and beyond the contributions already made by the other variables. The variable with the second largest F-to-remove is considered as the second most important, and so on.

For the recommended discriminating variables in D.Fn = V, the computed values for F-to-remove are as under:

<table>
<thead>
<tr>
<th>Variable</th>
<th>(X_1)</th>
<th>(X_2)</th>
<th>(X_6)</th>
<th>(X_9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-to-remove</td>
<td>8.8596</td>
<td>2.4616</td>
<td>0.0265</td>
<td>3.3719</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>(X_{10})</th>
<th>(X_{14})</th>
<th>(X_{16})</th>
<th>(X_{17})</th>
<th>(X_{18})</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-to-remove</td>
<td>1.4745</td>
<td>4.5267</td>
<td>0.1412</td>
<td>5.8003</td>
<td>1.2065</td>
</tr>
</tbody>
</table>

These values suggest \(X_1, X_{17}, X_{14}, X_9, X_2, X_{10}, X_{18}, X_{16}, X_6\) as the order of importance for individual characteristics with the highest rank getting assigned to \(X_1\) (Area Operated) and the lowest that to \(X_6\) (Number of Milch Cattle). Variables
$X_{17}$ (Off-farm urban employment), $X_{14}$ (Intensity of cultivation), and $X_9$ (Worker consumer ratio) occupy the second, third and fourth place, respectively.

Such a ranking is useful, no doubt, but conventionally the same objective is attained by calculating the percentage contribution of individual characteristics to the total distance. This procedure yielded 41.99 per cent discriminating power for area operated, 21.21 per cent for off-farm urban employment, 11.24 per cent for the intensity of cultivation, 11.20 per cent for the use of chemicals per acre and the remaining 14.36 per cent shared by the other five variables included in the function. These magnitudes are in conformity with the above mentioned ranking due to 'F-to-remove' statistic.

Curiously enough, the size of the operational holding turns out to be the most important discriminator but, fortunately, does not happen to be the only one. Consequently, total reliance on land holding categories for identifying the rural poor is definitely a misplaced one. Apart from size of the holding, several other characteristics need to be included in any appropriate 'target-group' approach seeking group identification in the first instance.

Briefly, we can say that, it is possible to sufficiently discriminate between poor and non-poor cultivating households, on the basis of easily available characteristics, mentioned
earlier. Further, of these nine characteristics, understandably, size of the holding is the dominant discriminator but other variables cannot be dismissed as aberrations. And, finally, the estimated function can be legitimately used for classification of unknown cases with a fair degree of predictive accuracy.
In drawing this list, we have been guided by the following: Heady (1964, op.cit.); Chayanov (1966, op.cit.); Coffey, Joseph D., "Personal Distribution of Farmers' Income by Source and Region", AJAE, Dec., 1968; Weeks, John, 'Uncertainty, Risk, and Wealth and Income Distribution in Peasant Agriculture', JDS, Oct., 1970; Shanin (1972, op.cit.), and Berry, Albert, 'Farm Size Distribution, Income Distribution and the Efficiency of Agricultural Production: Colombia', AER, May, 1972.

What constitutes healthy performance in agriculture is a moot point. Consequently, several traditions have emerged in the literature. Profit maximization (or the underlying concepts of 'technical efficiency' and 'price efficiency'), 'utility maximization', 'survivability in the market economy' or 'maintaining the labour consumption-pressure balance', to name a few, have been suggested as the main motivations of agricultural organizations. Under the circumstances, the term 'performance' appears to be too complex..... not amenable to any single quantitative formulation. Therefore, we have included the various constituents of income, productivity, expansion of land-base, and debt in this group.

Also refer to note 12, Chapter 4 for further readings and Singh, Ajit, Take-overs, Cambridge University Press, Cambridge, 1977, pp.6-10, for a brief review.


That u's are normally distributed with $E(u_i)=0$ and $E(u_i,u_j)=\sigma^2$ ; i=j $\sigma^2=0$ otherwise. For details about this model one can refer to any standard text-book, see, for instance, Kmenta (1971, op.cit.), pp.347-391.

On this point, refer to note 15, Chapter 4.
Given a large base, even lower returns (within limits) can produce higher aggregates. So in this sense the bigger is better. In addition, the new technology also bestowed a relative advantage upon the bigger farms, hence, strengthening their position further. See, Heady and others (1971); Rao (1975, op. cit.) pp. 135-151; Rudra (1982, op. cit.), Ch. X, and Jairath, Jaseen, 'Social Conditioning of Technology Use: A Study of Irrigation and Production in Punjab, 1965-70', EPW (ROA), March 29, 1986.

Urban influence variable or city-gravity variable, defined as the ratio of population of the nearest town (a proxy for size) and distance of the village from that urban centre, has been included to capture the influence of urbanization on the activities carried out on the farm. This formulation is due to Vandeventer, Lonnie Ray, An Economic Analysis of the Western Oklahoma Agricultural Land Market, Ph.D. Thesis, Oklahoma State University, 1979.

It is not easy to empirically work out the consequences of these local-power configurations. But interaction with the local population gives an impression that big landlords have cornered the input markets and have monopolized the credit institutions. As a result, the smaller farmers have to bid very high, may it be land, labour or credit. Thus, the relatively powerful have an edg. Myrdal, Gunnar in his The Challenge of World Poverty - A World Anti-Poverty Programme in Outline, Penguin Books, Middlesex, 1970, pp. 70-79 touches upon the relation between inequality and power. However, for detailed discussion on these issues, see, Chaudhuri, B.B, 'Rural Power Structure and Agricultural Productivity in Eastern India, 1757-1947'; Rudra, Ashok, 'Local Power and Farm-Level Decision Making', and Chakravarty, Sukhamoy, 'Power Structure and Agricultural Productivity' in Desai, M., Rudolph, S.H, and Rudra, A. (eds.) Agrarian Power and Agricultural Productivity in South Asia, Oxford University Press, Delhi, 1984.

Starting with Schultz, T.W., 'Investing in Poor People', AER, May, 1965 and—'Public Approaches to Minimize Poverty' in Fishman, L. (ed.) Poverty Amid Affluence, Yale University Press, New Haven, 1966, down to India's Seventh Five Year Plan 1985-90, Planning Commission, Govt. of India, New Delhi, 1985, much has been said about the positive role of education in the lot of the poor. But the Plan Document (ibid., p. 5) reports 'Enrolment targets have been exceeded, but a high rate...
of drop-outs means that the actual rate of attendance is very much lower*. Does that not imply that only the relatively better-off in India can afford to keep their children in schools? A study, got conducted by the author, strongly supported this view.

In the Laissez-faire system, poverty was assumed to be due to individual-specific factors such as incompetence, imprudence and laziness. However, in the welfare state, the persistence of poverty is attributed mainly to factors extraneous to and beyond the control of the poor. For an illuminating discussion, see Tendulkar, Suresh D., 'Economic Inequality in an Indian Perspective' in Beteille, Andre (ed.) Equality and Inequality, Theory and Practice, Oxford University Press, Delhi, 1983, pp.71-128.

This finding is very much in line with the current thinking on poverty and consequent remedial measures. For instance, Chopra, Pran in his article 'The Basis of Hope' and Tendulkar, Suresh D., in his 'Over View' in Population Poverty and Hope, Centre for Policy Research, New Delhi, 1983 come out with four suggestions for the removal of poverty as follows: (1) generating additional wage employment at higher wage rate for the wage-dependent poor households, (2) generating additional self-employment by raising the productivity of the existing asset bases of the poor households, (3) transferring new and/or existing assets to the assetless poor households in order to create the necessary self-employment opportunities, and (4) elimination of exploitative institutional arrangements that keep the incomes of the ultimate producers at a low level. Earlier Rath & Dandekar (1971, op.cit.) and Kurien, C.T., in his Poverty, Planning and Social Transformation, Allied Publisher (for ICSSR), New Delhi, 1978 had come out with a similar diagnosis of the problem though the suggested methods of attacking it differed.

See, Study of Implementation of Integrated Rural Development Programme, NABARD, Bombay, 1984; Evaluation Report on Integrated Rural Development Programme, P2O, Planning Commission, Govt. of India, New Delhi, 1985,
We could get just one study viz., Singh, I.J. & Pandey, U.K., 'Discriminant Function Analysis of Small Farmers and Landless in India', JAE, May 1981, which comes closer to dealing with this problem.


The selection of the variables in each group was guided by economic theory and pragmatism required placement of highly correlated characteristics in separate sub-groups. An alternative procedure would have been to feed all the variables simultaneously and fix more demanding selection criteria but we refrained from excessive reliance on empiricism.

A revised version (due to Dr. Murli Dhar, J.N.U., New Delhi) of BMD07M package was used for the purpose.


c.f., Klecka (1980, ibid.) p.46.

c.f., Klecka (1980, ibid.) p.47.

Since definite information about these prior probabilities was not available from the official publications, therefore, we depended on our own estimates of poverty ratios for the sampled population. Refer to Tables 5.2 & 5.3 for estimates of poverty.

(τ) (tau) is considered to be a direct measure of predictive accuracy and the most intuitive measure of discrimination. See, Klecka (1980, op.cit.) pp.50-51.

Because of the highly debated 'size and productivity' relationship [see, Bhardwaj (1974); Chadha (1978) and Saini (1979)]; and multicollinearity problem [see, Heady & Dillon (1960)]; Prof. O.P. Bagai & Prof. R.N. Soni (P.U. Chandigarh) and several other participants in All India Advanced Level Seminar in Probability, Statistical Inference and Related Fields, April 15-19, 1985 held at Chandigarh, advised the author against it.


Treating equally likely group membership for these households, the value of τ turned out to be 0.86, which means that classification based on the discriminating variables made 86 per cent fewer errors than would be expected by a random assignment.


Percentage contribution of the jth variable is given by
\[ \frac{\lambda_j d_j}{D^2_p} \times 100, \]

where \( \lambda_j \) = the jth discriminant coefficient;

\[ d_j = \bar{x}_{1j} - \bar{x}_{2j}; \quad (j=1,2,\ldots,p) \]

and \( D^2_p \) = the Mahalanobis' distance.

This approach, for calculating the percentage contribution of individual characteristics to the total distance measured, is quite popular with the researchers. For some applications and a good exposition of this method, see, Pandy & Singh (1981, op.cit.), and George et.al. (1984, op.cit.).