CHAPTER III
MATERIAL AND METHOD
MATERIAL AND METHOD

The present study is based on the two phases of field work conducted among Chenchus of Kurnool (C-K) and Mahaboobnagar district (C-M) of Andhra Pradesh. During the year 1977, a pilot survey was undertaken from July to August in both the district Chenchus and data on genetical demography and anthropometric measurements were taken. Again, in 1981, a four months field work was undertaken from March to July and data on genetical demography, Dermatoglyphics, blood groups, Taste sensitivity to Phenyl thiocarbamide, and colour blindness were collected.

3.1 Genetical demography

Genetical demography data have been collected from 18 settlements of Achampet taluq of Mahaboobnagar district comprising 504 families and 9 settlements from Atmakur and Nandyal taluqs of Kurnool district comprising 215 families (Table 13). Detailed bilateral three generation pedigrees for each couple was drawn. The demographic information was on their age, sex, marital status, relation with head of the family and age at marriage. The information from married women was on their menarcheal age, age at first birth, outcome of each pregnancy, number of children, their age, sex and marital status. Age information has been corroborated with pedigree information and also with many check points like major events and cross checking with elders. Marital
<table>
<thead>
<tr>
<th>Settlement</th>
<th>Number of households</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kurnool District</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Bairlutty</td>
<td>30</td>
<td>417</td>
</tr>
<tr>
<td>2 Nagalutty</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>3 Srisailam</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>4 Peddacheruvu</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>5 Basapuram</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>6 Mehanandi</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>7 Ahobilam</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>8 Sunnipenta</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>9 Abbersa je gunta</td>
<td>45</td>
<td>198</td>
</tr>
<tr>
<td>Mahaboobnagar District</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Appalapally</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>2 Chanchu guudem</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>3 Banala</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>4 Siddapur</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>5 Billakal</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>6 Vankteswarla Bevi</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>7 Macheram</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>8 Vatverlapally</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>9 Ferehabed</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>10 Udimmilla</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>11 Meradugu</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>12 Jangamreddy pally</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>13 Kollam, Kommanpenta</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>14 Thatigundal</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>15 Sarla pally</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>16 Pedra</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>17 Mannanoor</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>18 Ambagiri</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Samples drawn from different settlements/villages for various variables in the two districts

<table>
<thead>
<tr>
<th>Village / Settlement</th>
<th>ABC system</th>
<th>MN system</th>
<th>Rh system</th>
<th>PTC taste test</th>
<th>Colour blindness</th>
<th>Dermatoglyphics</th>
<th>Anthropometry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mahanagaram</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Appal pally</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>-</td>
<td>15</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>2 Chench gudem</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>-</td>
<td>21</td>
<td>25</td>
<td>-</td>
</tr>
<tr>
<td>3 Farshabad</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>24</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>4 Udhamila</td>
<td>25</td>
<td>-</td>
<td>25</td>
<td>-</td>
<td>32</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>5 Manmanoor</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>64</td>
<td>64</td>
<td>43</td>
<td>-</td>
</tr>
<tr>
<td>6 Jangamreddy pally</td>
<td>23</td>
<td>-</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>7 Gudibanda</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8 Rangaper</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9 Vatwora pally</td>
<td>30</td>
<td>26</td>
<td>30</td>
<td>16</td>
<td>42</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>10 Siddapur</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>31</td>
<td>31</td>
<td>-</td>
</tr>
<tr>
<td>11 Pedra</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>12 Sarla pally</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>13 Komman penta</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>189</td>
<td>63</td>
<td>135</td>
<td>111</td>
<td>229</td>
<td>121</td>
<td>54</td>
</tr>
<tr>
<td><strong>Kurnool District</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Srisailam</td>
<td>31</td>
<td>18</td>
<td>31</td>
<td>30</td>
<td>30</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>2 Haleluty</td>
<td>59</td>
<td>51</td>
<td>50</td>
<td>22</td>
<td>45</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>3 Rahenadi</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>27</td>
<td>24</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>4 Ahobilam</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>37</td>
<td>38</td>
<td>31</td>
<td>5</td>
</tr>
<tr>
<td>5 Faddachauru</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>-</td>
<td>40</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>176</td>
<td>103</td>
<td>145</td>
<td>116</td>
<td>177</td>
<td>146</td>
<td>49</td>
</tr>
</tbody>
</table>
FIG: 1
KURNOOL AND MAHABOOBNAGAR DISTRICTS OF A.P. SHOWING CHENCHU SETTLEMENTS
distance data have been collected from the married couple based on their travel route in Kilometers.

Demographic data were analysed as per the standard techniques following Barclay (1958).

3.1.1 Sex ratio: The sex ratio was expressed as the number of males per hundred females.

3.1.2 Fertility: Fertility was expressed in terms of the number of pregnancy terminations and live born off-spring per woman and per fertile woman. The fertile women are those who conceived at least once at the time of investigation.

3.1.3 The number of surviving off-spring was defined as the number of off-spring surviving to a given woman.

3.1.4 Mortality: Mortality was expressed both as number per woman and proportion over total pregnancies which includes both pre- and post-natal deaths.

(A) Pre-natal deaths: These include all foetal deaths before birth which include both abortions and still births.

(i) Abortions: Foetal deaths upto and including 6th month of gestation,
(ii) Still births: Foetal deaths from the start of 7th month of gestation to before birth.

(B) Post-natal deaths: These include all deaths from birth to the age of fifteen years. These are also referred as pre reproductive deaths and further
classified as 1. Infant deaths (deaths within one year), 2. Juvenile deaths (deaths before actual start of reproduction, i.e. fifteen years).

3.1.5 Sterility: Sterility has been defined as the percentage of women who have not conceived even once before the onset of their menopause.

3.1.6 Population growth rate: The growth rate of Chenchu population was determined assuming an exponential growth by the formula -

\[ \log_e \left( \frac{p_2}{p_1} \right) \]

where \[ r = \frac{\log_e \left( \frac{p_2}{p_1} \right)}{n} \times 100 \]

\[ n = \text{Number of years between census.} \]

3.1.7 Index of opportunity for selection (I):

The index of opportunity for selection (I), with its components \( \text{Im} \) and \( \text{If} \), has been calculated, following Crow (1958) formula as follows -

\[ I = \text{Im} + \frac{\text{If}}{\text{Ps}} \]

where

\[ I = \text{Index of opportunity for selection} \]

\[ \text{If} = \frac{\text{Vf}^2}{\text{X}} \]

\[ \text{Variance of the number of live born} \]

\[ = \frac{\text{Square of the mean number of live born}}{\text{Ps}} \]

\[ \text{Im} = \frac{\text{Id}}{\text{Ps}} \]

\[ \text{Proportion of children who died below 15 years of age} \]

\[ = \frac{\text{Proportion of survivors upto 15 years of age}}{\text{Ps}} \]
Pe = 1 - Pd

Im = Index of selection due to mortality component.

If = Index of selection due to fertility component.

3.1.8 Effective population size (Ne):

Wright's formula following Lasker (1952) has been used.

\[ Ne = \frac{\bar{N}K}{\sqrt{(N-1)\sigma^2 + NK(\bar{K} - 1)}} \]

where

\( N \) = Number of parents

\( \bar{K} \) = Mean number of offspring per couple

\( \sigma^2 \) = Variance of \( \bar{K} \)

3.1.9 Genetic drift: random drift per generation in the population has been calculated by the formula (Li, 1963),

\[ \sigma^2 q = \frac{q(1-q)}{2 Ne} \]

Where

\( \sigma^2 q \) = Variance due to drift

\( q \) = gene frequency

\( Ne \) = Effective population size

3.1.10 Co-efficient of breeding isolation: Index of isolation was calculated following Lasker (1960), which is the product of effective population size (Ne) and migration rate (m). Where (m) is the percent of population born elsewhere.
3.1.11 Inbreeding coefficient (autosomal):

Consanguinity data, collected through large pedigrees, has been calculated for Uncle-Niece, first cousins, first cousins once removed, second cousins and beyond as per the standard F values of 1/8, 1/16, 1/32 and 1/64 respectively. The weighted average of all inbreeding coefficient of the progeny including those with F = 0 provides the mean inbreeding coefficient (\( \alpha \)) of a population, the weights (\( F_i \)) being the relative frequency of the progeny with inbreeding coefficient (\( F_i \)). Therefore \( \alpha = \sum F_i F_i \).

3.1.12 Genetic load: Inbreeding effects on abortions, stillbirths, child mortality were examined through an exponential model (Morton et al., 1956)

\[ P_i = 1 - \exp \left( - (\alpha + \beta F_i) \right) \]

where

\( P_i = \) expected proportion of survivors

\( F_i = \) weights coefficient of inbreeding

The estimates \( \alpha \) and \( \beta \) (\( A \) and \( B \)) were obtained through weighted least square technique (Smith, 1967) and goodness of fit tested using appropriate distributions. \( h \) is an estimate of random mating load, and \( \sigma \) is an estimate of inbreeding load.
3.2 Anthropometry

Anthropometric data on 103 male Chenchu adults were collected from those of age group ranging from 18 to 64 from the two districts. 15 somatometric measurements on head, face and body were obtained from these individuals following Martin and Seller (1957, 1962) with anthropometric rod, sliding caliper, spreading caliper and a weighing machine. The settlements and the individuals covered are shown in table 4. The measurements taken on head are 1. Maximum head length, 2. Maximum head breadth, 3. Minimum frontal breadth, and 4. Horizontal circumference of head. The measurements on face are 1. Bigonial breadth, 2. Bisygomatic breadth, 3. Nasal breadth, 4. Nasal height, 5. Nasal depth, 6. Total facial height, and 7. Upper facial height. The measurements taken on body are 1. Stature, 2. Height acromion, 3. Chest girth, and 4. Weight. The measurements taken between the landmarks along with the instrument used are as follows:

3.2.1.1 Maximum head length: It is the distance between glabella and opisthocranion, i.e. the most projecting point on the dorsal surface of the head in the mid sagittal plane. Instrument used is spreading caliper.

3.2.1.2 Maximum head breadth: It is the distance between the two eurya points. Instrument used is spreading caliper.
3.2.1.3 Minimum frontal breadth: It is the distance between the two frontotemporalis. Instrument used is Spreading caliper.

3.2.1.4 Horizontal circumference of the head: It is the maximum circumference of the head taken horizontally. Instrument used is tape.

3.2.1.5 Bisygomatic breadth: It is the distance between the two zygia, i.e. the most lateral points on the zygomatic arch. Instrument used is Spreading caliper.

3.2.1.6 Biconial breadth: It is the distance between the two gonial points. Instrument used is Spreading caliper.

3.2.1.7 Nasal height: It is the distance between nasion and projective point of the nose. Instrument used is Sliding caliper.

3.2.1.8 Nasal breadth: It is the distance between the two alar, i.e. the most laterally placed points on the nasal wings. Instrument used is Sliding caliper.

3.2.1.9 Nasal depth: It is the distance between sub nasale and projective point of the nose. Instrument used is Sliding caliper.

3.2.1.10 Total facial height: It is the distance between nasion and gnathion. Instrument used is Sliding caliper.
3.2.1.11 **Upper facial height**: It is the distance between nasion and prosthion. Instrument used is Sliding caliper.

3.2.1.12 **Stature (Height vertex)**: It is the vertical distance from vertex to floor. Instrument used is anthropometric rod.

3.2.1.13 **Height acromion**: It is the vertical distance from acromion to the floor. Instrument used is Anthropometer.

3.2.1.14 **Chest girth**: It is the circumference of the chest at the level of nipples when a subject is breathing normally. Instrument used is a tape.

3.2.1.15 **Height**: Body weight taken by a standard weighing machine with the subject wearing minimum clothes.

Altogether 10 indices were derived out of these 15 measurements along with the standard formulae as follows.

<table>
<thead>
<tr>
<th>3.2.2.1 Cephalic index:</th>
<th>[\frac{\text{Head breadth}}{\text{Head length}} \times 100]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.2.2 Nasal index:</td>
<td>[\frac{\text{Nasal breadth}}{\text{Nasal length}} \times 100]</td>
</tr>
<tr>
<td>3.2.2.3 Total facial index:</td>
<td>[\frac{\text{Total facial height}}{\text{Bzygomatic breadth}} \times 100]</td>
</tr>
<tr>
<td>3.2.2.4 Upper facial index:</td>
<td>[\frac{\text{Upper facial height}}{\text{Bzygomatic breadth}} \times 100]</td>
</tr>
</tbody>
</table>
3.2.2.5  Transfronto parietal index:
Minimum frontal breadth

3.2.2.6  Gonic-zygomatic index:
Bignorial breadth

3.2.2.7  Frontozygomatic index:
Minimum frontal breadth

3.2.2.8  Transphelo facial index:
Bizygomatic breadth

3.2.2.9  Korperfuelle index:
(Rohre’s index)
Weight in grams

3.2.2.10  Fignets index:
Height vertex - Chest girth + Weight

3.3  Dermatoglyphics

Dermatoglyphics data is based on 256 Chenchu male individuals from both the districts (Kurnool district:146
Mahaboobnagar district:121). Impression of both the palms and
rolled prints of fingers have been collected employing ink
and pad method on white bond papers from unrelated individuals.
The age of the individuals in Kurnool district is from 5 to 50
years, whereas, in Mahaboobnagar district the ages vary
between 6 to 45 years. The prints have been analysed after
the methods of Cummins and Midlo (1961). Number of triradii
on palms were determined after Penrose (1968) and ridge
counting has been done as described by Holt (1968).
3.4 Serology

Blood was collected by finger pricking from each individual with the help of a sterilized pricking needle in 0.9% normal saline solution and 2% cell suspension was prepared for blood grouping. For all the blood grouping tests respective controls were always tested for checking the sera before starting the analysis.

3.4.1 A1 A2 O blood groups: 365 unrelated individuals were tested from both the districts (Kurnool: 176, Mahaboobnagar: 189) using high titre anti-A, Anti-B, and group O sera supplied by Haffkine Institute, Bombay. The moist chamber method (Soorman and Dodd, 1957) was followed. Group A blood specimens were subsequently tested for A₁ and A₂ utilising anti-A₁ serum (lectin) supplied by Daedalus Corporation, Bombay.

3.4.2 MN blood groups: For MN blood groups, 166 unrelated individuals from both the districts (Kurnool: 103, Mahaboobnagar: 63) were tested utilising anti-M and anti-N sera supplied by Gamma Biologicals, U.S.A. The micro tube technique as suggested by Race and Sanger (1968) was employed for detecting M and N antigens.

3.4.3 Rh(D) blood group: 280 unrelated individuals were tested for Rh(D) blood group utilising Rh(D) serum (human) supplied by Haffkine Institute, Bombay. The tests for Rh(D) were conducted by Albumin replacement technique.
3.5.1. **PTC tests:** Harris and Kalmus (1949) sorting technique was used for testing taste ability to Phenyl-thio-<br>ceramide (PTC) on 227 males and females of both the districts (Kurnool: 116, Mahaboobnagar: 111). 14 serial dilutions of Phenyl-thio-<br>ceramide salt were prepared with distilled water. The stock solution contained 0.13 gm of PTC in 100 ml of distilled water and No. 2 and No. 3 solutions are 0.0659 gm/100 ml and 0.0325 gm/100 ml respectively. Each subsequent dilution had half the concentration of PTC.

Following Das (1956), about 800 ml of distilled water was taken in a pyrex glass beaker and heated in a water bath with the temperature maintaining at 60° to 65°. PTC crystals weighing 1.3 gms were added to the solvent in the beaker and continuously stirred until the entire substance dissolved and a clear solution was obtained. The solution was then poured in a measuring flask and its volume was increased to one litre by adding distilled water. Each subject was given 5 ml of the solution at once. The threshold number was noted after confirming it with the help of sorting technique of Harris and Kalmus (1949).

3.5.2. **Colour blindness:** 406 unrelated individuals from both the districts were tested for colour blindness data with the help of Ishihara colour plates (1968) to diagnose the Protanopes and deuteranopes.
3.6 Selection of sample

A house to house survey was conducted for demography in each settlement of Chanchu village. The Chenchu settlements are known as gudem (locally) in Telugu language and also by Chenchus which means collection of a number of households. Each settlement varies in the number of households ranging from 5 to 50. Smaller settlements were lumped with the bigger settlements for certain working facility. Most of the settlements in the upper plateau and lower plateau of Khammam taluq of Mahaboobnagar district and most of the settlements in Anmakur taluq of Kurnool district were covered for genetical demography data. No particular sampling procedure could be applied to the collection of data for Serology, FTC, Colour blindness, Dermatoglyphics and Anthropometric data as is often the case with anthropological field work. Exploratory nature of investigation, representing from the main concentration settlements might have yielded the proportional sample. Samples collected represents various settlements from men and women of different ages and also from schools.

3.7 Calculation of gene frequencies

3.7.1 $A_1A_280$ system: The frequencies of $F_1\cdot P_2Q$ and $r$ were estimated computing from the following formulae (Mourant et al., 1976).

\[ P_1 = 1 - \sqrt{\bar{u} + \bar{a}_2 + \bar{b} + \bar{a}_2\bar{b}} \]
\[ P_2 = \sqrt{\bar{u} + \bar{a}_2 + \bar{b} + \bar{a}_2\bar{b}} - \sqrt{\bar{u} + \bar{b}} \]
Under genetic equilibrium $q$ should be unity that never appears exact in practice. Therefore the corrected values $p_1^t$, $p_2^t$, $q^t$, $r^t$ were obtained making simple adjustments as follows.

$$p_1^t = p_1 \left(1 + \frac{0}{2}\right) \quad p_2^t = p_2 \left(1 + \frac{0}{2}\right)\quad q^t = q \left(1 + \frac{0}{2}\right) \quad r^t = (r + \frac{0}{2}) \left(1 + \frac{0}{2}\right)$$

These estimates closely approximate with the maximum likelihood estimates. Since these form a set of solution to the ML equations, they are fully efficient and their variances may be taken as those of the ML estimates.

To obtain the variance for these estimates an information matrix is generated based on the following table:

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Probability $F$</th>
<th>$dF(p_1)$</th>
<th>$dF(p_2)$</th>
<th>$dF(q)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$</td>
<td>$r^2$</td>
<td>$-2r$</td>
<td>$-2r$</td>
<td>$-2r$</td>
</tr>
<tr>
<td>$s$</td>
<td>$q^2 + 2qr$</td>
<td>$-2q$</td>
<td>$-2q$</td>
<td>$2r$</td>
</tr>
<tr>
<td>$A_1$</td>
<td>$p_1^2 + 2p_1p_2 + 2p_1r$</td>
<td>$2p_2 + 2r$</td>
<td>$0$</td>
<td>$2p_1$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$p_2^2 + 2p_2r$</td>
<td>$-2p_2$</td>
<td>$2r$</td>
<td>$-2p_2$</td>
</tr>
<tr>
<td>$A_1B$</td>
<td>$2p_1q$</td>
<td>$2q$</td>
<td>$0$</td>
<td>$2p_1$</td>
</tr>
<tr>
<td>$A_2B$</td>
<td>$2p_2q$</td>
<td>$0$</td>
<td>$2q$</td>
<td>$2p_2$</td>
</tr>
</tbody>
</table>
Since \( p_1, p_2 \) and \( q \) are codominant over 0, it is suffice to generate a 3x3 information matrix and after inverting to get the variances for these an extended Covariance matrix whose rows and columns add up to 1 will be used for obtaining the variance of \( V(r) \). Square rooting of these variances give the standard errors (Balakrishnan, 1980).

3.7.2 MN blood group frequencies: MN blood group frequencies were obtained by the following formula (gene counting method).

\[
\begin{align*}
    m &= \frac{2 \, MN + MN}{2} \\
    n &= \frac{MN + 2 \, MM}{2}
\end{align*}
\]

The variances were obtained as:

\[
V(m) = \frac{mn}{2 \, G} \quad V(n) = \frac{mn}{2 \, G}
\]

3.7.3 Rh(D) blood group: Rh(D) blood group frequencies were obtained by the formula

\[
\begin{align*}
    d &= \sqrt{dd} \\
    D &= 1 - d^2 \\
    V(d) &= \frac{1 - d^2}{4 \, G}
\end{align*}
\]

3.7.4 PTC taste gene frequencies: Let \( O \) and \( A \) be the proportions of the PTC tasters and non-tasters in a sample of \( G \). Let \( 'T' \) and \( 't' \) be the frequencies of the corresponding alleles.

\[
\begin{align*}
    t &= \sqrt{R} \\
    T &= 1 - t
\end{align*}
\]
Variances of the alleles are obtained as

\[ V(t) = \frac{1 - t^2}{4} = V(T) \]

are square root of variances.

3.7.5 Goodness of fit tests:

(a) the goodness of fit for ABO system was done by the Karipearson's method as follows.

\[
\frac{1}{N} \chi^2 = \frac{\sum (O_i - E_i)^2}{r^2} + \frac{\sum (A_i - B_i)^2}{\rho^2 + 2pr} + \frac{\sum (\rho^2 + 2qr)}{q^2 + 2qr} + \frac{\sum (\rho^2 + 2qr)}{2pq}
\]

where \( N \) = total sample size. \( \rho, q, r \) = corrected gene frequencies. If the \( \chi^2 \) value lies below 6, then the said sample is supposed to be a good fit.

(b) Homogeneity test to find out the excess or deficiency of ABO blood group under Hardy–Weinberg equilibrium was computed by the \( O/E \) statistic

\[ D = 1 - \rho + q + r \]

\[ \sigma = \sqrt{\frac{pq}{2N}} \]

Where \( D \) = 1 - \( \rho + q + r \)

(c) For other genetical systems the following formula was utilised

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

where \( O_i = \text{Observed frequency} \)

\( E_i = \text{Expected frequency of the } i^{th} \text{ phenotype} \)

\( k = \text{total number of phenotype} \)

The degree of freedom depends on the number of alleles.
3.8 The following standard statistical formulas were utilised.

3.8.1 Arithmetic Mean (\( \bar{X} \)) : It is one of a measure of central tendency.

\[
\bar{X} = A + \frac{\sum f_i d_i}{N} X C
\]

where, 
- \( A \) = arbitrary assumed mean
- \( f_i \) = frequency
- \( d_i \) = deviation from assumed mean
- \( N \) = total sample
- \( C \) = class interval

3.8.2 Standard Deviation (\( \sigma \)) : It is a measure of dispersion. It is the square root of variance.

\[
\sigma = \sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2} X C
\]

where,
- \( f_i d_i \) = Total of frequency deviation
- \( N \) = Total sample
- \( C \) = Class interval

3.8.3 Variance (\( \sigma^2 \)) : It is the square of the standard deviation. It is the arithmetic mean of the squared deviations from the arithmetic mean.

\[
\sigma^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 \quad \text{where } x_i = \text{individual observation} \quad \bar{X} = \text{mean}
\]

or

\[
\sigma^2 = \frac{\sum f_i d_i^2}{N} \frac{(\sum f_i d_i)^2}{\sum f_i} X C
\]
3.8.4 Coefficient of variation (C.V.) : It measures the degree or percent of variability in the character relative to the average of the group.

\[ C.V. = \frac{\sigma}{\mu} \times 100 \]

3.8.5 Standard error (S.E.) : It is calculated for various statistical constants as

\[ S.E. \text{ of mean} = \frac{\sigma}{\sqrt{N}} \quad S.E. \text{ of variance} = \frac{\sigma}{\sqrt{2N}} \]

\[ S.E. \text{ of S.D.} = \frac{\sigma}{\sqrt{2N}} \quad S.E. \text{ of C.V.} = \frac{C.V.}{\sqrt{2N}} \]

3.8.6 Correlation Coefficient (r):

A coefficient measuring the degree to which points in a cartesian diagram tend to fall near a straight line. It is an analysis of covariation of two variables.

\[ r = \frac{1}{n} \sum xy - \overline{xy} \]

or

\[ r = \sqrt{\frac{\sum dx \cdot dy}{\overline{dx} \cdot \overline{dy}} - \left( \frac{\sum dx}{n} \right)^2 - \left( \frac{\sum dy}{n} \right)^2} \]

\[ S.E. \text{ of } r = \frac{1 - r^2}{\sqrt{n}} \]
3.8.7 Significance tests:

(a) $\chi^2$ test: It is a test of significance used to determine if a set of observed frequencies ($O_i$) differs from those of expected frequencies ($E_i$) for a set of variables distributed in a contingency table.

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

The degrees of freedom is equal to the product of row - 1 x column - 1 number.

(b) 't' test: It is used to find out the significant difference between the two mean values of quantitative variables using the formula

$$t = \frac{M_1 - M_2}{\sqrt{S^2 + S^2_2}}$$

where, $M_1$ and $SE_1$ are Mean and Standard error of one variable, whereas, $M_2$ and $SE_2$ are Mean and Standard errors of the second variable.

3.8.6 Skewness and Kurtosis: These are the two tests which are used to test the deviations from the normal distribution. Skewness is measured by $g_1$ statistic

$$g_1 = K_3 / K_2 \left( \sqrt{K_2} \right)$$

where, $K_3 = \text{Average of the third power of deviations from mean}$

$K_2 = \text{Variance.}$
If $g_1 = 0$ symmetry in the population is demonstrated.
Positive $g_1$ indicates an excess of items smaller than mean, whereas, negative $g_1$ indicates slight asymmetry with excess of items larger than the mean drawing the peak of the frequency curve towards right.

Kurtosis is measured by $g_2$, a statistic based on the sum of the fourth powers of deviations from mean

$$g_2 = \frac{K_4}{K_2^2}$$

where, $K_4$ = sum of the fourth powers of deviations

If $g_2 = 0$ then there is no departure from normality. A positive value of $g_2$ indicates an excess of items near the mean and far from it. Negative values of $g_2$ results from flat topped distribution curves.

3.9 Genetic distance

Genetic distance is a concept which is utilised to measure the genetic difference between populations. It is calculated from gene frequency data for a number of genetic loci. Using the matrix of genetic distances between populations one can construct phylogenetic or evolutionary tree and estimate the time lapsed since the separation of any two populations. Besides this, genetic distance is used to measure the degree of micro evolution among local populations and sometimes it can be related to geographic, linguistic
differences or historical pattern of migrations of populations and thereby one can study the mechanism of microevolution. Genetic distance can also be utilised to study the proportion of admixture.

There are various genetic distance methods proposed by many scientific workers, which can be broadly grouped into two categories. 1. Indices based on statistical considerations, and 2. Indices based on biological consideration.

1. Indices based on statistical consideration:

These indices are designated to provide a measure of genetic relationship between two populations based on differences in allelic frequencies, taking into account the statistical properties of allelic frequencies. No assumptions are made regarding how the genetic relationship if any has come about. The essential feature of all these indices is representing the populations by points in a multi-dimensional space and measuring distance geometrically.

Bhattacharya's angular distance between multinomial populations (1946), Mahalanobis' design to measure on the basis of variation in a large number of quantitative characters (Mahalanobis, 1936; Rao, 1952), Sanghvi's genetic distance (1953) analogous to the chi-square statistic, Edwards and Cavalli-Sforza's (1964) E, which is an improvement of the Bhattacharya's distance. Edwards E, which is an improvement of the Edwards and Cavalli-Sforza's distance and Balakrishnan & Sanghvi, E and G (1968) falls in this category.
2. Indices based on biological consideration:

The basic idea here is that of kinship defined by Malecot. The coefficient of kinship is the probability of identity of descent of two genes, chosen at random either from the same population or from two different populations.

Nei's distance of Maximum (D₁), Standard (D) and Minimum (Dₘ) and the index of Morton which is based on right's hierarchical model are genetic distances in this category.

The merits and demerits of those different distance measures were thoroughly discussed by Roychoudhury (1975) and Balakrishnan (1981) and it is also found that there is a high correlation between these various distance measures.

For the present work, Balakrishnan and Sanghvi's (1968) distance $D^2$ is utilised, because, with the help of this measure distance can be calculated for genetic markers, anthropometric measurements, dermatoglyphic characters separately and an augmented distance pooling all these various category of characters can be obtained. $D^2$ like Mahalonobis generalised distance ($D^2$) makes use of the actual dispersion matrices of individual populations by combining them in a common dispersion matrix based on their weighted values. One more advantage of this measure is that the dispersion matrix for calculation of genetic distance is derived directly from the dispersion matrix obtained during the maximum likelihood estimation of gene frequencies.
This distance which takes into account of the variances and covariances of the allele frequencies can be computed as follows

\[ G^2 = \sum_{j=1}^{s-1} \sum_{k=1}^{s-1} C_{jk} d(m,n,j) d(m,n,k) \]

Where

\[ C_{jk} = (C_{jk})^{-1}, \quad C_{jk} = \sum_{j=1}^{q} \frac{H^2(i) \delta(i,j,k)}{\sum_{j=1}^{q} N(i)} \]

where, \( s \) = number of alleles at a particular locus

\[ d(m,n,j) = \left[ p(m,j) - p(n,j) \right]^2 \]

\[ p(m,j) = \text{frequency of allele } j \text{ in population } m \]

\[ p(n,j) = \text{frequency of allele } j \text{ in population } n \]

\[ \delta(i,j,k) \] is the \((j,k)\)th element in the dispersion matrix of the allele frequencies in population \( i \) based on a sample of \( N(i) \). The dispersion matrix is being obtained during the maximum likelihood estimation of the allele frequencies.

Information from different loci is pooled by simple summation of the squared distance.

\[ q = \text{number of populations} \]

3.10 Clustering

After getting the pair-wise distances among a set of populations, we have to get some pattern from the distance matrix and then to assess these patterns in relation to geographic, linguistic, ecological and other relevant factors.
This is being done by grouping related populations known as cluster analysis. There are various methods to get this clusters (Sokal and Sneath, 1963), and for the present work clustering by binary split using the method of fusion (Galskriahnan et al., 1975) was utilised. The method of fusion starts with 2 clusters, one consisting of the 2 populations showing the smallest distance and the other consisting of rest. Each of the other populations are added one by one to the first cluster and removed from the other cluster. The cluster of 3 populations which gives the largest distance becomes the basis for the next stage of getting a cluster of 4 populations. The process is continued until one cluster consists of only one population.

3.11 Dendrogram

The clustering process can be represented in 2 dimensions by a dendrogram. A dendrogram is a tree-like figure lying on its side to facilitate listing of the populations against the final stems. Two stems will join at a level given by distance between the two clusters represented by the stem. The scale on the abscissa will start with 0 on the right. The end points of the final stems are usually equally spaced. The ordinate can be given a meaning if clustering has been done by a binary split method. The correlation coefficient between the original distances and the phanetic distances will give an idea about the representativeness of the dendrogram.
3.12 **Principal co-ordinate analysis**

The dendrogram tells only at what level 2 stems come together and it can not represent relationships among the populations adequately. The populations can be represented as points using a rectangular co-ordinate system in 1, 2 or 3 dimensions with minimum loss of information by the method of principal co-ordinate analysis (Gower, 1972). The method adopted is based on principal components of the between product sum matrix using mean values of the transformed variates used in obtaining distances. An approximate 3 dimensional representation is obtained as suggested by Zenghui et al., (1971).