UNIVERSITY OF CAPE COAST

# APPLICATION OF DISCRIMINANT ANALYSIS FOR EVALUATING THE STATUS OF CORONARY HEART DISEASE RISK

BY

TETYANA MEDVYEDYEVA

DISSERTATION SUBMITED TO THE DEPARTMENT OF MATHEMATICS AND STATISTICS OF THE SCHOOL OF PHYSICAL SCIENCES, FACULTY OF SCIENCE, UNIVERSITY OF CAPE COAST IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN STATISTICS

MAY 2007

## CANDIDATE'S AND SUPERVISOR'S DECLARATION

### CANDIDATE'S DECLARATION

I hereby declare that this dissertation is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere

Candidate's Signature

Candidate's Name Tetyana Medvedyeva

Date 8 - 2 - 08

### SUPERVISOR'S DECLARATION

I hereby declare that the preparation and presentation of the dissertation were supervised in accordance with the guidance on supervision of dissertation laid down by the University of Cape Coast

Supervisor's Signature

Supervisor's Name PROF. NICHOLAS N. N. NSOWAH-NUAMAH

Date

# ACKNOWLEDGEMENTS

I am indebted to my supervisor, Prof N Nuamah Nsowah, the Government Statistician, for his help, suggestions and guidance in writing this dissertation

I am grateful to Prof B K Gordor of University of Cape Coast, for his interesting lectures, which eventually triggered my interest in multivariate methods and consequently gave me the idea for the topic of this dissertation

My gratitude also goes to Dr Natasha Mensah, Head of Department, Mathematics and Statistics, University of Cape Coast, for her great encouragement and support in diverse ways

I am also thankful to Mr Patrick Dickar, a Senior Biochemist at Korle-Bu Teaching Hospital, who helped me to collect the needed data and also served as the resource person for some medical facts that I used in this dissertation

To all other people who contributed in one way or the other in writing this dissertation I say thank you

# TABLE OF CONTENTS

# CHAPTER 1

## INTRODUCTION

### BACKGROUND OF THE STUDY

Traditionally, being healthy means being absent of illness. If someone did not have a disease, then he or she was considered healthy. The overall health of a nation or specific population was determined by numbers measuring illness, disease, and death rates. Today this rather negative view of assessing individual health and health in general is changing. A healthy person is one who is not only free from disease but also fully well.

Being well, or wellness, involves the interrelationship of many dimensions of health: physical, emotional, social, mental, and spiritual. This multifaceted view of health reflects a holistic approach, which includes individuals taking responsibility for their own well-being (Zilva and Pannall, 1984). Every year millions of people world-wide are falling ill and die of diseases that can be prevented. Our health and longevity are affected by various choices we make every day. Medical reports tell us that we abstain from smoking, drugs, excessive alcohol consumption, fat and cholesterol, and if we get regular exercise our rate of disease and disability will significantly decrease. These call for positive change in the lifestyle of people.

Health education and statutory measures help to improve health-related behaviour by the entire population and are fundamental aspects of prevention of many diseases. The appropriate health delivery and effective preventive strategies against diseases can save the lives of millions of people worldwide.

## STATEMENT OF THE PROBLEM

Diseases can be congenital and hereditary or can be acquired during a person's life span Some of the acquired diseases are due to a person's life style Knowledge and understanding of the causes of acquired diseases may help in changing the habits and behaviors of people, which would in effect lead to the prevention of these diseases in society One of the acquired diseases that is on the increase world-wide is Coronary Heart Disease (CHD) which can be prevented through appropriate life style

The research problem of this dissertation is to discriminate the individuals categorized as high-CHD risk group from relatively healthier individuals categorized as low-CHD risk group and to determine factors that significantly differentiate between these two groups

## STUDY OBJECTIVES

The study aimed to show the application of discriminant analysis for evaluating the status of coronary heart disease risk The specific objectives were

1 Explore the factors that are likely to be responsible for coronary heart disease risk

2 Develop a discriminant function, which will best discriminate between the high-CHD risk individuals and low-CHD risk individuals

3 Examine statistical and practical significance of the developed discriminant function

4 Determine the factors that contribute most to the development of high coronary heart disease risk

5 Assess the validity of discriminant analysis

6. Classify the individuals with unknown status of coronary heart disease risk into high risk group or low risk group based on the values of predictor variables

## LITERATURE REVIEW

Coronary heart disease (CHD), also called coronary artery disease, is a condition that affects arteries that deliver blood, oxygen and nutrients to the heart. CHD results when a fibrous tissue, called plaque, builds up inside the arterial walls, causing a partial or complete blockage of blood flow. This condition is commonly called arteriosclerosis. When the heart is deprived of oxygen-carrying blood, a sensation of pain, burning, pressure or other type of discomfort, which is generally felt in the chest, may occur. When the heart's supply of oxygen is completely cut off by a blood clot, then heart attack occurs, resulting in permanent tissue death for part of the heart (http://www.taskforce.com).

Coronary heart disease is the major cause of death and a common cause of illness in adults in most developed and many developing countries. The disease affects both young and old. Men initially have a greater risk for developing CHD than women. Once a woman reaches the age of 50, however, her risk for heart disease eventually equals or surpasses that of a man. Experts also confirm that the risk of CHD rate more than doubles every 10 years afte the age of 55 in both males and females (Zilva and Pannall, 1984).

There is no clear statistics on the state of coronary heart disease in Africa, and particularly in Ghana. The latest research shows that the death from heart attack caused by CHD in Africa is not as common as in Europe and North America (www.wrongdiagnosis.com/c/coronary-heart-sisease/stats-country.htm). This however does not mean that CHD is not a threat for African population. The uncommonness of death from CHD is explained by the relatively short life expectancy due to the high level of death from infections such as HIV, malaria and tuberculosis.

Doctors in most African countries indicate that the intake of fat food has risen in the past few years and the intake of fiber-containing foods has fallen. The level of obesity (especially in females) has risen enormously, physical activity of urban population decreased to the minimum and prevalence of hypertension exceeds that in the white population. With these increases in risk factors, medical experts expect urban Africans to attain high mortality rate from CHD, particularly as the mortality rate from infectious diseases go down, the mortality from CHD will increase at an alarming rate (www.wrongdiagnosis.com/coronary-heart-sisease/stats-country.htm)

**Symptoms of Coronary Heart Diseases**

Someone affected with coronary heart disease (CHD) may have a wide range of symptoms. Some people with CHD do not have any symptoms at all and consequently may get no warning that they have a disease, until they have a heart attack. Others may have mild chest discomfort or shortness of breath, while some with CHD have frequent chest pain that interferes with their daily activities.

The classic symptom associated with CHD is called angina. Angina is usually described as a chest discomfort that can be experienced as a "pressure," "squeezing," "burning" or "heaviness" sensation in the center of the chest, underneath the rib cage. The discomfort can sometimes spread to the neck, jaw or left arm. Other symptoms that may accompany angina include sweating, nausea and shortness of breath (http://www.taskforce.com )

Angina is sometimes classified into two types: stable and unstable based on the occurrence of symptoms. Stable angina is usually somewhat predictable. It may occur with increased exertion or extreme emotion, or following a large meal. The symptoms do not last long (usually one to five minutes) and are relieved by rest or by taking nitroglycerin (http://www.taskforce.com). Unstable angina is less predictable. It may occur with little exertion or during rest, it often comes and goes at frequent intervals, and may be accompanied by more severe symptoms.

### Risk factors for Coronary Heart Disease

Risk factors for CHD are circumstances or conditions that increase the likelihood of developing this disease. Coronary heart disease risk factors are divided into two groups: controllable and uncontrollable. High blood pressure (hypertension), high blood cholesterol, smoking, obesity, physical inactivity, diabetes, and abnormal lipid profile measurements are regarded as controllable risk factors. Other risk factors such as heredity, age, gender (men are at greater risk initially) are considered as uncontrollable. In rarer instances, CHD can result from other medical conditions. Some examples include: the formation of a blood clot in the coronary artery due to an abnormal blood clotting condition, inflammation of the coronary arteries due to an autoimmune disorder, spasms of the coronary arteries from cocaine abuse, congenital abnormalities of the coronary arteries.

A growing literature has documented associations between several psychosocial factors and increased incidence of CHD. The most prominent among these are stress, lack of social support, depression, and socioeconomic status (http://www.taskforce.com).

Many deaths associated with heart disease are preventable. Experts agree that the decline in death rates from heart disease is possible only if the public adopt a healthier lifestyle. Many aspects of lifestyles are the key to prevention of coronary heart disease. By making healthier lifestyle choices, we hold the key to lowering our risk for developing and preventing illness and death from CHD. Some recommendations that can help to lower the risk of coronary heart disease can be seen in Appendix B1

## DATA COLLECTION

The research work employs secondary data, collected from the Central Laboratory of Korle-Bu Teaching Hospital, Accra. Data mainly consist of a group of tests, known as blood lipid profile, which are often ordered together to determine risk of CHD. The tests have been shown to be good indicators of whether someone is likely to have a heart attack in the near future or not

The data was collected between the months of August and September 2006. Convenience sampling was used to draw a sample of 250 patients referred to the laboratory. The sample consists of two parts, drawn at two different time occasions. The first 100 individuals with recorded low-CHD risk and the first 100 individuals with recorded high-CHD risk were included in the sample on the first occasion, making up a total of 200 patients with known CHD risk status. The results of these 200 patients were used for the preliminary analysis as well as for the further analysis - to develop and validate the discriminant function which best separates the individuals who are at high risk of developing coronary heart disease (CHD) from relatively healthier individuals with low CHD risk. On the second occasion, the data from 50 patients with unknown status of CHD risk were collected and added to the 200 patients with known CHD risk status, making up a total sample of 250 patients. The 50 unknown CHD risk cases were used for predictive classification of cases: the developed discriminant function was used to classify individuals with unknown CHD risk into one of the groups (high or low CHD risk) based on the values of predictor variables. The collected data can be seen in the first ten columns of Appendix A.

Since the data collected for this study was over a period of two months, care was taken not to collect multiple results of the same patients who were being monitored (on medications) and had to report to the laboratory on several occasions for the same tests. Also all necessary precautionary measures were taken or observed by qualified staff so as to generate reliable and accurate laboratory results. The details of precautionary measures taken can be seen in Appendix B2.

### Variables description

The data for discriminant function analysis consists of one criterion (dependent) variable and seven predictor (independent) variables. The criterion variable - CHD risk consists of two mutually exclusive and collectively exhaustive groups: group 1 - low-CHD risk and group 2 - high-CHD risk. The predictor variables are: age, body mass index (BMS), cholesterol level, triglycerides level, direct high density lipoprotein (D-HDL), low density lipoprotein (LDL) and fasting blood sugar level (FBS). Another variable - gender was considered in the preliminary analysis.

**Body mass index (BMI)** is calculated by dividing weight in kilogram (kg) by height in meters squared (Zilva and Pannall, 1984).

**Table 1.1 Body Mass Index WHO classification** (http://www.healthguidance.org/How-to-Prevent-Coronary-Heart-Disease-and-Heart-Attack.html.)

| Body Mass Index (BMI) | WHO classification | Popular description |
|---|---|---|
| <18.5 $kg/m^2$ | Underweight | Thin |
| 18.5-24.9 $kg/m^2$ | normal weight | "healthy", "acceptable" weight |
| 25.0-29.9 $kg/m^2$ | grade 1 overweight | Overweight |
| 30.0-39.9 $kg/m^2$ | grade 2 overweight | Obesity |
| >40.0 $kg/m^2$ | grade 3 overweight | morbid obesity |

In the study, weights and heights of the patients were collected and based on their values, BMI was calculated

The other four variables cholesterol, triglycerides, direct high-density lipoprotein (D-HDL) and low density lipoprotein (LDL) serum parameters constitute the lipid profile tests and usually done together to ascertain the risk of coronary heart disease

**Cholesterol** and **triglycerides** are the major circulating lipids in the bloodstream Cholesterol is utilized by the cells of the body for the syntheses and repair of membranes Cholesterol is also present in structures of the brain and the central nervous system It is also the precursor of various steroid hormones such as progesterone, testosterone, estrogen and others that play major roles in human reproductive process The normal range for cholesterol level in serum is [3.64 - 6.40] mmol/L (Pagana and Pagana, 2003) Triglycerides serve as an energy source that can be stored as fat in adipose tissues and burned as fuel by muscle and other tissues The normal range for triglycerides level in serum is [0.30 - 1.70] mmol/l

(Pagana and Pagana, 2003). Low levels of cholesterol and triglycerides may be associated with malnutrition whereas high levels with excess intake of fatty food.

**Direct high-density lipoprotein (D-HDL)** and **low density lipoprotein (LDL)** are major lipoproteins in the bloodstream. Their function is to transport lipids from one site in the body to another. The purpose of D-HDL is to remove cholesterol from the blood vessels. It is known as "good" or "healthy" cholesterol. The normal range for D-HDL level is [1.07 - 3.00] mmol/L (Pagana, 2003). The purpose of LDL is to transport cholesterol to cells. The normal range for LDL level is [0 – 156] mg/dL (Pagana, 2003). Low levels of D-HDL and high levels of LDL contribute to the development of coronary heart disease.

**Fasting blood sugar (FBS)** test determines level of glucose in the blood. The normal range for FBS level is [3.40 - 6.40] mmol/L (Pagana, 2003). High levels of blood sugar can cause diabetes, which is one of the recognized risk factors for coronary heart disease.

**CHD risk** is assessed by the cholesterol-to-D-HDL ratio. According to the National Institutes of Health the ratio values within the interval [1.0 - 4.0] is considered as relatively low CHD risk, ratio values greater than 4.0 indicate that the person is at risk of getting CHD (Pagana, 2003). In the study work individuals with the ratio values within [1.0 - 4.0] were put in group 1 - low-CHD risk and individuals with the ratio values greater than 4.0 were put in group 2 - high-CHD risk.

## LIMITATIONS

Due to time and logistics constrains sample size of only 250 patients was considered The data was collected within a short period of two months August and September 2006 The source of data was also restricted to only one hospital - Korle-Bu Teaching Hospital, Accra The CHD risk factors (predictor variables) used in this dissertation were the ones, which were easily assessable There are other well-known CHD risk factors such as blood pressure, hereditary factors, behavioral factors (smoking and alcoholic abuse), stress, and status in society, which were not considered in this study

Constraints mentioned above were also the reason for the fact that majority of the patients in the sample were 50 years old and above (these were the patients referred to the laboratory by their physicians) This fact may be the reason why the age was found to be a weak factor for the development of coronary heart disease

## OUTLINE OF DISSERTATION

This dissertation has five chapters The first chapter is an introduction to the study It discusses the aspect of health of a person and the effect of lifestyle of a person on his or her health It also stresses the importance of health education, which can help to improve health related behaviours that can lead to the prevention of many diseases Chapter one further presents the research problem as well as the objectives of the study The chapter also gives literature review on coronary heart disease (CHD), looks at the risk factors for CHD and the methods of prevention of CHD

The second chapter explains the statistical tools and methods used to analyse the data The chapter gives a brief background of two-group discriminant function analysis and its applications It also looks at the two different purposes and procedures for conducting discriminant analysis discriminant predictive analysis and discriminant classification analysis

Chapter three presents the preliminary analyses of the study These include basic characteristics of the patients as well as the comparison of results of low-CHD risk group patients and high-CHD risk group patients The chapter also examines the relationship between body mass index and measured parameters of lipid profile together with the dependence of CHD risk ratio on body mass index and fasting blood sugar levels

Chapter four presents further analyses. It shows the application of discriminant analysis for evaluating the status of coronary heart disease by the means of developing a discriminant function that separates the individuals who are at high risk of coronary heart disease from the relatively healthier individuals with low CHD risk Chapter four also shows how the developed function can be used to predict whether an individual is likely to have a heart attack in the near future or not

The final chapter, chapter five, concludes the report by discussing the results and the outcomes found by the study It also gives recommendations based on the results of the study and raises a few issues, which may be the subject for further research

# CHAPTER 2

# REVIEW OF METHODS

The dissertation employs two statistical methods: descriptive statistics and two-group discriminant function analysis. Statistical tools, such as tables and graphs were also used for pictorial representation of the data collected and for summarizing the obtained study results.

## TWO-GROUP DISCRIMINANT FUNCTION ANALYSIS

Discriminant function analysis is a multivariate technique concerned with separating distinct sets of observations and with allocating new observations to previously defined groups. Discriminant analysis method used to analyse data when there are several, usually continuous, independent variables and one categorical dependent variable.

There are two very distinct purposes and procedures for conducting discriminant analysis. The first procedure, discriminant predictive analysis, is used to derive discriminant functions from a set of weighted independent variables. The two-group discriminant analysis model involves linear combinations of the following form:

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

where

$D$ = discriminant score

$b_0, b_1, b_2, b_3, \dots, b_k$ = discriminant coefficients or weights

$X_1, X_2, X_3, \dots, X_k$ = predictor or independent variables

The coefficients, or weights ($b_i$), are estimated so that the groups differ as much as possible on the values of the discriminant function (Malhotra, 2004)

The second procedure, discriminant classification analysis, uses the predictive functions derived in the first procedure to either classify initial set of the data of known group membership, thereby validating the predictive function, or to classify new sets of observations of unknown group membership

## DISCRIMINANT ANALYSIS FOR PREDICTION

Discriminant analysis conducted for predictive purposes uses an initial data set with known group membership to derive the Discriminant function that can be used to predict group membership of future observations

Let $\boldsymbol{X}$ be a $\boldsymbol{kx1}$ vector of $\boldsymbol{k}$ independent variables. $X' = (X_1, X_2, X_3, \ldots, X_k)$ The observed values of $\boldsymbol{X}$ differ to some extent from one class to the other The variance-covariance matrix of $\boldsymbol{X}$ is given by $\sum$ Let $\gamma$ be a $\boldsymbol{kx1}$ vector of weights associated with the $\boldsymbol{k}$ variables $\boldsymbol{X}$

The discriminant function will be given by

$$\xi = X'\gamma \qquad (1)$$

The sum of squares for the resulting discriminant scores will be given by

$$\xi'\xi = (X'\gamma)'(X'\gamma) = \gamma' X X' \gamma = \gamma' T \gamma \qquad (2)$$

where $T = X'X$ is the total sum of squares and cross products ($SSCP_t$) matrix for the $\boldsymbol{k}$ variables. Since

$$SSCP_t = SSCP_b + SSCP_w \qquad \text{or} \qquad T = B + W,$$

where $\boldsymbol{B}$ and $\boldsymbol{W}$ are, respectively, between-groups and within-group SSCP matrices for the $\boldsymbol{k}$ variables, (2) can be written as

$$\xi'\xi = \gamma'(B + W)\gamma = \gamma'B\gamma + \gamma'W\gamma \tag{3}$$

where $\gamma'B\gamma$ and $\gamma'W\gamma$ are, respectively, the between-groups and within-group sum of squares for the discriminant score $\boldsymbol{\xi}$.

The objective of discriminant analysis is to estimate the weight vector, $\gamma$, of the discriminant function given by (1) such that

$$\lambda = \frac{\gamma'B\gamma}{\gamma'W\gamma} \tag{4}$$

is maximized. The vector of weights, $\gamma$, can be obtained by differentiating $\lambda$ with respect to $\gamma$, and equating to zero:

$$\frac{\partial\lambda}{\partial\gamma} = \frac{2(B\gamma)(\gamma'W\gamma) - 2(\gamma'B\gamma)(W\gamma)}{(\gamma'W\gamma)^2}$$

Dividing through by $\gamma'W\gamma$,

$$\frac{2(B\gamma - \lambda W\gamma)}{\gamma'W\gamma} = 0$$

$$(B - \lambda W)\gamma = 0$$

$$(W^{-1}B - \lambda I)\gamma = 0 \tag{5}$$

Equation (5) is a system of homogeneous equations and for a nontrivial solution

$$|W^{-1}B - \lambda I| = 0 \qquad (6)$$

It can be shown that for two groups, ***B*** is equal to

$$B = \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)(\mu_1 - \mu_2)' = C(\mu_1 - \mu_2)(\mu_1 - \mu_2)' \qquad (7)$$

where $\mu_1$ and $\mu_2$, respectively are ***px1*** vectors of means for group 1 and group 2. $n_1$ and $n_2$ are number of observations in group 1 and group 2, respectively, and $C = n_1 n_2/(n_1 + n_2)$ is a constant

Therefore, (6) can be written as

$$[W^{-1}C(\mu_1 - \mu_2)(\mu_1 - \mu_2)' - \lambda I]\gamma = 0$$

$$CW^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)'\gamma = \lambda\gamma$$

$$\frac{C}{\lambda} = [W^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)'\gamma] = \gamma \qquad (8)$$

Since $(\mu_1 - \mu_2)'\gamma$ is a scalar, (8) can be written as

$$\gamma = KW^{-1}(\mu_1 - \mu_2) \qquad (9)$$

where $K = C(\mu_1 - \mu_2)'\gamma/\lambda$ is a constant. Since the within-group variance-covariance matrix, $\sum_w$, is proportional to ***W*** and it is assumed that $\sum_1 = \sum_2 = \sum_w = \sum$, (9) can be written as

$$\gamma = K\sum^{-1}(\mu_1 - \mu_2) \qquad (10)$$

Assuming K = 1,

$$\gamma = \sum^{-1}(\mu_1 - \mu_2) \quad \text{or} \quad \gamma' = (\mu_1 - \mu_2)'\sum^{-1} \qquad (11)$$

Different values of the constant $K$ give different values for $\gamma$ and therefore the absolute weights of the discriminant function are not unique, only the ratio of the weights is unique (Sharma,1996)

Once the vector of weights, $\gamma$ is known, the prediction of the discriminant score $\xi$ is routine, since all values in the formula (1) are known

## DISCRIMINANT CLASSIFICATION ANALYSIS

Discriminant classification analysis, uses the predictive function derived in the discriminant predictive analysis to either classify initial sets of data of known group membership, thereby validating the predictive function or if the function has been validated to classify new sets of observations of unknown group membership The classification of the initial data set is but an extension of the predictive discriminant analysis in that, the predictive discriminant scores form the basis of the decision rule used to classify this same set of objects into the two groups

In contrast to the initial data set, where group membership i known the same decision rule may be applied to other sets of data However, when we classify data sets other than the initial set from which the predictive analysis was conducted, we are no longer engaged in predictive discriminant analysis, but rather in discriminant classification analysis A number of methods are available for classifying sample and future observations but in this study the cutoff-value and statistical decision methods were used for developing classification rules

**Cutoff-Value Method**

Cutoff classification method involves the partitioning of the discriminant or the variable space into two mutually exclusive regions - $R_1$ and $R_2$. The value of the discriminant score that divides the space into the two regions is called the cutoff value. The classification problem reduces to determining a cutoff value that divides the space into $R_1$ and $R_2$ regions. The value selected to be a cutoff value is the one that minimizes the number of incorrect classifications or misclassification errors. A commonly used cutoff value that minimizes the number of incorrect classifications for the sample data is

$$C = \frac{\bar{Z}_1 + \bar{Z}_2}{2},$$

where $\bar{Z}_1$ and $\bar{Z}_2$ is the average discriminant scores for group 1 and group 2. The formula assumes equal sample sizes for the two groups.

For unequal sample sizes, the cutoff value is given by

$$C = \frac{n_1 \bar{Z}_1 + n_2 \bar{Z}_2}{n_1 + n_2},$$

where $n_1$ and $n_2$ are the number of observations in group 1 and group 2.

Furthermore, any given case will be classified into group 1 if its value of discriminant score is less than the cutoff value and into group 2 if its value of discriminant score is greater than the cutoff value (Sharma,1996).

In classifying observations, two types of errors can occur. An observation coming from group 1 can be misclassified into group 2. Let C(2/1) be the cost of this misclassification

Similarly, an observation coming from group 2 can be misclassified into group1. Let $C(1|2)$ be the cost of this misclassification. Naturally, we would like to use the criterion that minimizes misclassification costs. The statistical decision theory method for developing classification rules is discussed below.

**Statistical Decision Theory**

Consider the case where we have only one discriminating variable, **X**. Let $\pi_1$ and $\pi_2$ represent the populations for the two groups and $f_1(x)$ and $f_2(x)$ be respective probability density functions for the random vector, **X**. Let $C$ be the cutoff value.

The conditional probability of correctly classifying observations in group 1 is given by

$$P(1|1) = \int f_1(x)dx$$

The conditional probability of correctly classifying observations in group 2 is given by

$$P(2|2) = \int f_2(x)dx$$

where $P(i|j)$ is the conditional probability of classifying observations to group $i$ given that they belong to group $j$.

The conditional probability of misclassification is given by

$$P(1|2) = \int f_2(x)dx \quad \text{and} \quad P(2|1) = \int f_1(x)dx \qquad (12)$$

Assuming that the prior probabilities of a given observation belonging to group 1 and group 2, respectively, are $p_1$ and $p_2$, the various classification probabilities are given by

$P$(correctly classified into $\pi_1$) = $P(1|1)\, p_1$

$P$(correctly classified into $\pi_2$) = $P(2|2)\, p_2$

P(misclassified into $\pi_1$) = $P(1|2)\, p_2$

P(misclassified into $\pi_2$) = $P(2|1)\, p_1$

The total probability of misclassification (TPM) is given by

$$TPM = P(1|2)\, p_2 + P(2|1)\, p_1 \tag{13}$$

The total cost of misclassification (TCM) is given by

$$TCM = C(1|2)\, P(1|2) + C(2|1)\, P(2|1)\, p_2 \tag{14}$$

were $C(i|j)$ is the cost of misclassifying into group $i$ an observation that belongs to group $j$

Classification rules can be obtained by minimizing either (13) or (14). In other words, the cutoff value $c$ should be chosen such that either TPM or TCM is minimized. If (13) is minimized then the resulting classification rule assumes equal costs of misclassification. If (14) is minimized then the resulting classification rule assumes unequal misclassification costs. The minimization of (14) results in a more general classification rule. It can be shown that the rule which minimizes (14) is given by the following

Assign an observation to $\pi_1$ if

$$\frac{f_1(x)}{f_2(x)} \geq \left[\frac{C(1|2)}{C(2|1)}\right]\left[\frac{p_2}{p_1}\right]$$

and assign to $\pi_2$ if

$$\frac{f_1(x)}{f_2(x)} < \left[\frac{C(1|2)}{C(2|1)}\right]\left[\frac{p_2}{p_1}\right]$$

## ASSUMPTION OF DISCRIMINANT ANALYSIS

The assumption of discriminant analysis is that the data come from a multivariate normal distribution and that the covariance matrices for the predictor variables are equal across groups (Sharma,1996)

The assumption of multivariate normality is necessary for the significance tests of the discriminator variables and the discriminant function. If the data do not come from a multivariate normal distribution then, in theory, none of the significance tests are valid Classification results, in theory, are also affected if the data do not come from a multivariate normal distribution

Discriminant analysis also assumes that the covariance matrices of variables are homogeneous across groups As with multivariate normality, violation of equal covariance matrices assumption affects the significance tests and the classification results

However, discriminant analysis is fairly robust to these assumptions The violations of the normality assumption are usually not "fatal" meaning, that the significance tests and the classification results are still "trustworthy" if the data is at least approximately normal Generally, if coefficient of skewness of the variable is within the range of -1 0 to 1 0 and kurtosis less than 2 0, the variable is considered approximately normal and is acceptable for running discriminant analysis (Harlow, 2005) Similarly, minor deviations from homogeneity of covariances are not that important, however, before running discriminant analysis it is advisable to review the within-groups variances and correlation matrices In particular a scatterplot matrix can be produced and can be very useful for this purpose

# CHAPTER 3

## PRELIMINARY ANALYSIS

### BASIC CHARACTERISTICS OF THE PATIENTS

Between the months of August and September 2006, the data on blood lipid profile and FBS of 200 patients were collected from the Central Laboratory of Korle-Bu Teaching Hospital, Accra. The examination of the data shows that the mean age of the patients was 52.61 years, as presented in Table 3.1. The youngest respondent was 20 years old and the oldest was 97 years.

**Table 3.1: Descriptive statistics of age of the patients**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Statistic |
| Age | 200 | 20 | 97 | 52.61 | 15.241 |

The age distribution of the patients referred to the laboratory (Fig 3.1 below) reveal that the age group (50 – 59) years old has the highest number (32%) of cases reported to the laboratory, followed by the age group (60 – 69) years old with 18.5% and (40 – 49) years, 16.5% of total cases. Young (20 – 29) years old and middle aged people (30 – 39) years old attracted 10.5% and 10% of total cases respectively, whilst old people (70 – 79) years old and (80 and over) had the low (9%) and the least (3.5%) number of reported cases respectively.

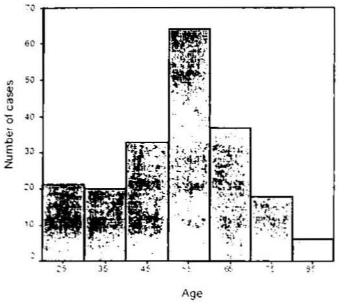


**Fig. 3.1 Age distribution of the patients referred to the laboratory**

More females (57% of the total recorded cases) reported to the hospital than males (43%). Appendix C However, as it can be seen from Fig 3.2, more males were found to be at high-CHD risk than females Out of all males reported to the hospital 54.7% were categorized as high-CHD risk group patients, compared to only 46.5% of females

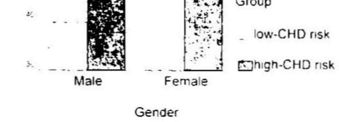


**Fig. 3.2: Distribution of males and females found to be at high and low CHD risk**

## COMPARISON OF RESULTS OF LOW-CHD RISK GROUP AND HIGH-CHD RISK GROUP PATIENTS

Analyses of data from two groups of patients reveal that the average age of patients in high-CHD risk group (57.5 years) is about ten years higher than in the low-CHD risk group (47.7 years). This information is displayed in Fig 3.3 below. This observation is in agreement with the well-recognized fact in medicine that incidence of CHD rises steeply starting at about age 50 years and the disease is most common at the age of 60 years (http://taskforce.com/)

Fig 3.4 compares body mass index (BMI) in two groups of patients. We can see that while the individuals in low-CHD risk group have normal or slightly overweight body mass, the individuals in the high-CHD risk group on the average are obese (BMI > 31kg/m$^2$)

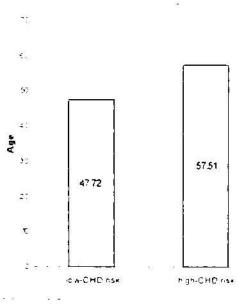


**Fig. 3.3: Average age of the patients referred to the laboratory**

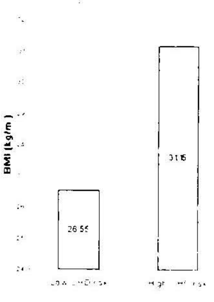


**Fig. 3.4: Average BMI of the patients referred to the laboratory**

From the Fig.3.5, we can observe that the average level of cholesterol in the low-CHD risk group is within the normal range of [3.64 - 6.4] mmol/L. On the other hand, the average level in the high-CHD risk group exceeds the normal range.

Fig 3.6 reveals that even though the average triglycerides level in the high-CHD risk group is within the normal range of [0.30 - 1.70] mmol/L, it is almost 1.5 times higher than in the low-CHD risk group.

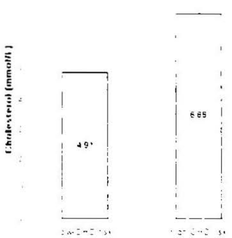


Fig. 3.5: Average cholesterol level

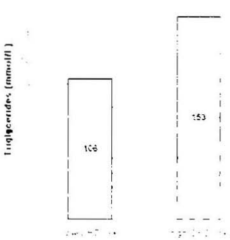


Fig. 3.6: Average triglycerides level

Generally, we can conclude from the two bar charts above that, the concentration of lipids in the bloodstream of patients categorized into high-CHD risk group is higher as compared to the low-CHD risk group patients.

Similarly, Fig 3.7 and Fig.3.8, representing the average levels of direct high-density lipoprotein (D-HDL) and low-density lipoprotein (LDL), depict a clear difference in the values for the low-CHD risk and high-CHD risk groups. Despite the fact that the average levels of D-HDL for both groups are within the acceptable range of [1.07 – 3.00] mmol/L, the level of D-HDL in the high-CHD risk group is noticeably lower than in the low-CHD risk group. This fact indicates that the ability of D-HDL to remove cholesterol from the blood vessels in high risk group is less than in the low risk group.

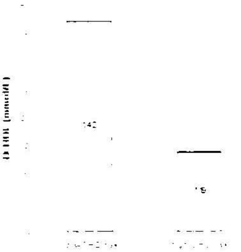


Fig. 3.7: Bar chart showing the average D-HDL level

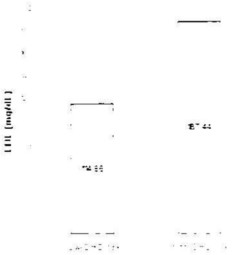


Fig. 3.8: Bar chart showing the average LDL level

The average levels of LDL show that while the value for low-CHD risk group is within the normal range of [0 – 156] mg/dL, the average value for the high-CHD risk group greatly exceeds the normal range. This indicates that LDL transports more cholesterol to the cells in high risk group patients as compared to the low group patients.

## RELATIONSHIP BETWEEN *BMI* AND MEASURED PARAMETERS OF LIPID PROFILE

Excessive weight, which can be brought about by the excess intake of fatty foods and carbohydrates, plays a major role in the function of body cells. The presence of these excess macromolecules in the body impedes the cell's normal metabolic activities. These may lead to different causes of diseases. Overweight (BMI ≥25 kg/m$^2$) and obesity (BMI ≥ 30kg/m$^2$) is not only a risk factor for CHD but is a cause for developing of other coronary heart disease factors, for example hypertension, hypercholesteroleamia, diabetes and others (http://www.healthguidance.org/How-to-Prevent-Coronary-Heart-Disease-and-Heart-Attack.html)

Fig 3.9 further supports the above mentioned fact. It clearly shows that the cholesterol level steadily increases with increase in BMI

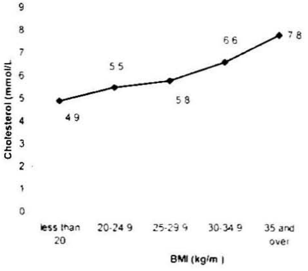


**Fig.3.9 Relationship between cholesterol level and BMI**

The average level of cholesterol rises steadily from 4.9 mmol/L for patients with the normal weight to 7.8 mmol/L for morbid obese patients

Fig 3.10 reveals that higher body mass index is associated with higher triglycerides level and lower D-HDL-cholesterol level

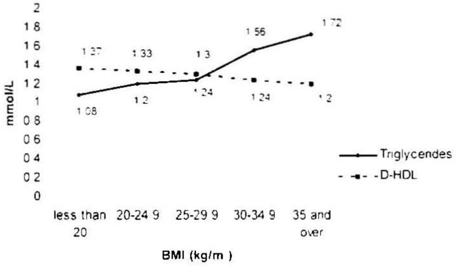


**Fig.3.10 Relationship between triglycerides, D-HDL and BMI**

The average level of triglycerides rises steadily from 1.08 mmol/L for patients with the normal weight to 1.72 mmol/L for obese patients. At the same time as the BMI values increase from normal to obese, the D-HDL level decreases from 1.37 mmol/L to 1.2 mmol/L respectively. The observations mentioned above indicate that as the weight of a person increases the amount of fat stored in the body cells also increases and the ability to remove fat from the cells decreases.

Both figures (Fig 3 7 and Fig3.8) clearly indicate the fact that excessive body mass affects the normal level of lipids in the blood and consequently can become a cause of hypercholesteroleamia and other risk conditions of coronary heart disease

## DEPENDENCE OF CHD RISK ON BODY MASS INDEX AND FASTING BLOOD SUGAR

Fig 3 11 shows a normal CDH risk ratio for patients with normal body weight However, as BMI increases from overweight (25-29 9) kg m$^2$ to severely obese (35 and over) kg m$^2$, the CHD risk ratio increases above the normal range from 4 6 to 6.6

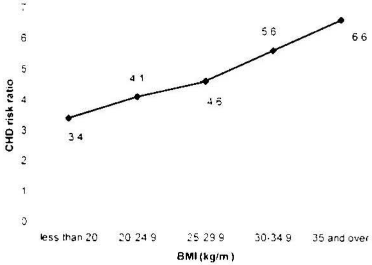


**Fig.3.11 Relationship between CHD risk ratio and BMI**

This suggests that body mass is an important factor responsible for the development of CHD We also can say that the patients with normal body mass are less likely to be at risk of CHD and the patients with excessive body mass are more likely to be at risk of CHD

Looking at Fig.3 12, we can observe that there is some influence of level of FBS on CHD risk ratio For the examined data, results show that irrespective of the level of FBS the average value for CHD risk ratio is out of the normal range This means that a sizable number of examined patients, belonging to the high-CHD risk group have normal levels of FBS However, patients with higher FBS level tend to have higher CHD risk ratio

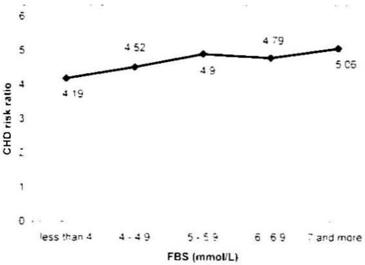


**Fig.3.12 Relationship between CHD risk ratio and BMI**

# CHAPTER 4

## FURTHER ANALYSES

### INTRODUCTION

As part of the study, the data were used to develop a discriminant function which best separates the individuals who are at high risk of developing coronary heart disease (CHD) from relatively healthier individuals with low CHD risk The discriminant analysis was also aimed at predicting whether an individual is likely to have a heart problem in the near future or not

The data for discriminant function analysis consists of one criterion (dependent) variable and seven predictor (independent) variables which are well-recognized factors of coronary heart disease The criterion variable - CHD risk, consists of two mutually exclusive and collectively exhaustive groups group 1 - low-CHD risk and group 2 - high-CHD risk The predictor variables are age, body mass index (BMI), serum cholesterol level, serum triglycerides level, serum direct high density lipoprotein (D-HDL), serum low density lipoprotein (LDL) and fasting blood sugar level (FBS) Based on these variables, the model for the discriminant function is of the form

*Group Risk  f(Age BMI Cholesterol Triglycerides D-HDL LDL FBS)*

The examined sample consists of 250 patients laboratory results, collected from the Central Laboratory of Korle-Bu Teaching Hospital, Accra Out of the total observed cases 100 patients were known to be at low-CHD risk and were categorized into group 1 - low-CHD risk group. Another 100 patients were known to be at high-CHD risk and were categorized into group 2 - high-CHD risk group. The other 50 patients were with unknown status of CHD risk. The first 200 patients with known status of CHD risk were randomly divided into two parts. The first part consists of 70% of the known risk cases (or 54% of the total cases) and was used as analysis sample for the estimation of the discriminant function. The other 30% of the known risk cases (or 26% of the total cases) served as holdout sample and were reserved to validate the discriminant function. The details of division of the cases with known status of coronary heart disease risk are described in Appendix C.

The patients with unknown risk status, representing 20% of the total sample, were used for predictive classification of cases. The results of discriminant analysis were used to classify individuals with unknown risk status into one of the groups (high or low CHD risk) based on the values of predictor variables. The full sample division can be seen in the Analysis Case Processing Summary Table, Appendix D.

## ASSUMPTIONS OF DISCRIMINANT ANALYSIS

Discriminant analysis assumes that data come from a multivariate normal distribution and that the covariance matrices for predictor variables are equal across groups. These assumptions were checked prior to performing discriminant analysis.

Multivariate normality assumption was verified by examining the skewness and kurtosis values of the predictor variables. Generally, a skewness value within the range of -1.0 to 1.0 and a kurtosis less than 2.0 indicate that the variable is approximately normal and are acceptable for running discriminant analysis (Harlow, 2005).

Table 4.1 provides descriptive statistics for the predictor variables. We can observe that all the skewness values are within the accepted range of normality, except the triglycerides, which slightly exceeded the acceptable limits. Kurtosis values also are all within the acceptable range of normality.

**Table 4.1: Descriptive statistics of patients parameters**

| | N | Minimum | Maximum | Mean | Std Dev | Skewnes | Kurtosis |
|---|---|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic |
| Age | 200 | 20 00 | 97 00 | 52 6150 | 15 34124 | - 060 | - 168 |
| BMI | 200 | 19 30 | 41 00 | 28 8480 | 4 73551 | 396 | - 410 |
| Cholesterol | 200 | 2 81 | 12 05 | 5 8935 | 1 57457 | 951 | 1 953 |
| Tnglycendes | 200 | 36 | 3 00 | 1 2903 | 60391 | 1 095 | 917 |
| D-HDL | 200 | 64 | 2 41 | 1 3062 | 26933 | 525 | 1 521 |
| LDL | 200 | 24 00 | 350 00 | 151 1500 | 56 85402 | 922 | 1 390 |
| FBS | 200 | 3 50 | 10 30 | 5 3375 | 1 03071 | 820 | 1 780 |

Even though the skewness for triglycerides is slightly out of the tolerable range, there does not appear to be enough nonnormality in the data, so we can assume that the data are approximately normally distributed.

The assumption of homogeneity of covariance matrices can be tested with Box's M Test, which tests null hypothesis of equal population covariance matrices. The results of the test which are displayed in the Appendix E1, indicate that the null hypothesis is rejected. However, this test is strongly influenced by even slight nonnormality of the data and may not be accurate in our case, since our data are only approximately normally distributed. Examining the matrix scatterplots of the two groups can better test the assumption of homogeneity of variance-covariance matrices for approximately normally distributed data (Leech *et al.* 2001). The scatterplots for group 1 and group 2 are presented in Fig 4.1 below.

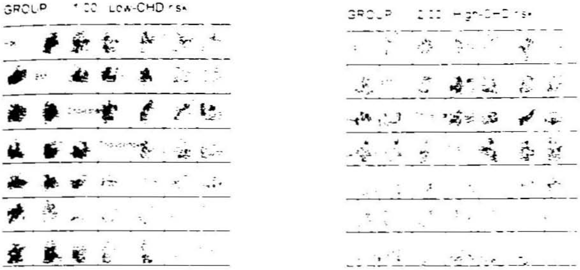


**Fig.4.1: Matrix scatterplots for two groups of patients**

The scatterplots for the same variables appear to be similar in variability for the two groups suggesting that the assumption of homogeneity of variance-covariance matrices is met.

The correlations between various pairs of predictor variables were also examined to ensure that variables are not very correlated (multicollinear). Pearson correlation table, Appendix E2 shows low to moderate correlations among the predictors except the [illegible] Cholesterol pair that has a high correlation of 0.886.

## DEVELOPMENT OF THE DISCRIMINANT FUNCTION

The analysis sample used for the estimation of the discriminant function [illegible] 135 cases (70% of all cases with known CHD risk status). [illegible] 68 [illegible] belong to the low-CHD risk group and 67 to the high-CHD risk group [illegible]

Table, Appendix F, provides basic descriptive statistics for each of the independent (predictor) variables for each outcome group separately, and for the whole analysis sample

The significance for discriminating variables was checked by testing of equality of group means, using Wilks'Λ as the test statistic The results of the test are displayed in Table 4 2

**Table 4.2: Test of Equality of Group Means**

| | Wilks' Lambda | F | df1 | df2 | Sig |
|---|---|---|---|---|---|
| Age | 923 | 11 061 | 1 | 133 | 001 |
| BMI | 703 | 56 113 | 1 | 133 | 000 |
| Cholesterol | 593 | 91 293 | 1 | 133 | 000 |
| Triglycerides | 831 | 27 131 | 1 | 133 | 000 |
| D-HDL | 833 | 26 669 | 1 | 133 | 000 |
| LDL | 600 | 88 569 | 1 | 133 | 000 |
| FBS | 971 | 4 026 | 1 | 133 | 047 |

We can observe that all seven variables in the discriminant model are significant predictors at significance level of 0 05, meaning that the selected discriminating variables significantly differentiate between the two groups Each variable potential is measured by the value of Wilks' Lambda Smaller values indicate the variable is better at discriminating between groups Table 4 2 suggests that cholesterol level is the best at discriminating between groups, followed by LDL level and BMI

Since the criterion variable consists of two groups, only one discriminant function is estimated The eigenvalue associated with this function (Table 4 3 below) is 1 503 and it accounts for 100 percent of explained variance

**Table 4.3: Eigenvalue**

| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
|---|---|---|---|---|
| 1 | 1.503 | 100.0 | 100.0 | .775 |

The practical significance of discriminant function is assessed by the square of canonical correlation associated with the function. From Table 4.3, square of canonical correlation is$(0.775)^2 \approx 0.6$. That is 60% of the variation between the two groups is accounted for by the discriminating variables.

The statistical significance of discriminant function was tested, based on Wilks'Λ test statistic. The result of the test, displayed in the Table 4.4, shows that Wilks'Λ associated with the function is 0.4, which transforms to a chi-square of 118.812 with 7 degree of freedom. This is significant beyond the 0.01 level.

**Table 4.4: Wilk's Lambda test of significance of discriminant function**

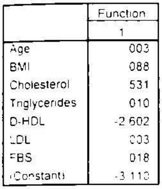
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---|---|---|---|---|
| 1 | .400 | 118.812 | 7 | .000 |

Standardized canonical function coefficients, displayed in Appendix G, indicate how heavily each variable is weighted in order to maximize discrimination of groups. The relative importance of the predictors can be seen in Structure Matrix in Appendix G. The structure correlations or discriminant loadings show the simple correlations between each predictor and the discriminant function and represent the variance that the predictor shares with the function. The variables with greater magnitude of structure correlation are considered to contribute more to the discriminating power of the function. Generally, variables with loadings greater or equal to |0.3| are considered as important predictors for discriminant function (Harlow, 2005). With this criterion, five out of seven variables in discriminant function (cholesterol, LDL, BMI, triglycerides and D-HDL) appear to be important predictors. From the Structure Matrix table it appears that cholesterol, LDL and BMI contribute most to the discriminating power of the function, with age and FBS contributing least.

We can also notice that the ordering in Structure Matrix is the same as that suggested by the test of equality of group means (Table 4.2) and is different from that in the Standardized Canonical Discriminant Function Coefficients, Appendix G. This disagreement is likely due to the collinearity between cholesterol and LDL levels noted in the correlation matrix. Since the Structure Matrix is unaffected by collinearity, it is safe to say that this collinearity has only inflated the importance of same variables in the Standardized Canonical Discriminant Function Coefficients Table.

Furthermore, the unstandardized discriminant function coefficients were estimated and presented in Table 4.5.

**Table 4.5: Unstandardized discriminant function coefficients**



| | Function |
|---|---|
| | 1 |
| Age | 003 |
| BMI | 088 |
| Cholesterol | 531 |
| Tnglycendes | 010 |
| D-HDL | -2 602 |
| LDL | 003 |
| FBS | 018 |
| (Constant) | -3 113 |

Based on the values of the coefficients we can form the discriminant function

$$D= -3.11+0.003(age)+0.088(BMI)+0.531(chol.)+0.01(trig.)-2.602(D\text{-}HDL)+0.003(LDL)+0.018(FSB)$$

This function best discriminates between two groups of patients low-CHD risk group and high-CHD risk group, and can be used for predictive classification of new cases. The values of the predictor variables for the patients with unknown status of coronary heart disease can be substituted into the discriminant function to calc 'ate the discriminant scores. Based on the value of the score, a patient can be assigned to either low-CHD risk group or to the high-CHD risk group.

## DISCRIMINANT CLASSIFICATION ANALYSIS

The developed discriminant function was then used to classify initial set of data with known group membership. The purpose of classification of observations of known grouping is merely to see how well the derived function predicts group membership using the subject data from which it was derived. In contrast to the classification of the initial data set, where group membership is known, the same decision rule may be applied to classify data with unknown group membership, thus the developed discriminant model can be used to classify cases with unknown status of CHD risk into low or high CHD risk groups

Prior to classifying the unknown cases, the values of discriminant function was examined. The group centroids, showing the value of the discriminant function calculated at the group means, were estimated and can be seen in Table 4.6

**Table 4.6: Groups Centroids**

| Group | Function<br>1 |
|---|---|
| Low-CHD risk | -1.228 |
| High-CHD risk | 1.228 |

The signs of the centroids suggest that higher values of age, cholesterol, BMI, triglyceride and FBS and lower values of HDLC are more likely to result in high-CHD risk.

The cutoff value or the value of discriminant score that divides the discriminant space into two regions (low-CHD risk and high-CHD risk) and minimizes the number of the misclassifications for the sample data is

$$Z_{cu} = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$$

were $\bar{Z}_1$ is the average discriminant score of group 1

$\bar{Z}_2$ is the average discriminant score of group 2

Since the average discriminant scores for group 1 and 2 are given [illegible] presented in the Table 4.7, the cutoff value is

$$C = \frac{-1\ 208 + 1\ 226}{2} = 0\ 009$$

Therefore, any given patient will be classified into low-CHD risk if the value of his discriminant score is less than the cutoff value and into high-CHD risk group if his discriminant score is greater than the cutoff value

The developed discriminant function was then applied to the raw values of the variables in the holdout sample consisting of 65 cases (30% of cases with known CHD status) for validation purpose The values of the predictor variables for the cases in the holdout sample were substituted into the function and the discriminant scores of holdout cases were calculated The cases were then assigned to the group whose centroid is the closest The calculated discriminant scores can be seen in column 13, Appendix A The function was also applied to the analysis sample to validate the original group cases Furthermore, the developed discriminant model was used to cross-validate the analysis sample classify each case in analysis sample, while leaving it out from the model calculations The percentage of cases correctly classified for each validating procedure was then calculated The classification results are displayed in Table 4 7 and they show how successful the developed discriminant function is in classifying the cases

**Table 4.7 Classification Results**

| | | | | Predicted Group Membership | | |
|---|---|---|---|---|---|---|
| | | | Group | Low-CHD risk | High-CHD risk | Total |
| Cases Selected | Original | Count | Low-CHD risk | 67 | 1 | 68 |
| | | | High-CHD risk | 9 | 58 | 67 |
| | | % | Low-CHD risk | 98.5 | 1.5 | 100.0 |
| | | | High-CHD risk | 13.4 | 86.6 | 100.0 |
| | Cross-validated | Count | Low-CHD risk | 66 | 2 | 68 |
| | | | High-CHD risk | 9 | 58 | 67 |
| | | % | Low-CHD risk | 97.1 | 2.9 | 100.0 |
| | | | High-CHD risk | 13.4 | 86.6 | 100.0 |
| Cases Not Selected | Original | Count | Low-CHD risk | 31 | 1 | 32 |
| | | | High-CHD risk | 8 | 25 | 33 |
| | | | Ungrouped cases | 27 | 23 | 50 |
| | | % | Low-CHD risk | 96.9 | 3.1 | 100.0 |
| | | | High-CHD risk | 24.2 | 75.8 | 100.0 |
| | | | Ungrouped cases | 54.0 | 46.0 | 100.0 |

The classification results based on the analysis sample indicate that (67 + 58)/135 = 0.962 or 96.2% of the cases are correctly classified. Cross-validation correctly classifies (66 + 58) = 0.919 or 91.9% of the cases. When the classification analysis is conducted on the independent holdout sample, we obtained (31 + 25)/65 = 0.862 or 86.2 % of the cases correctly classified. We also can observe that the discriminant analysis did better at predicting low-CHD risk group than high-CHD risk group. Given the account to the high percentage of correctly classified cases, the validity of discriminant analysis can be judged as satisfactory.

The developed discriminant function was also applied to classify the 50 patients with unknown status of CHD risk into low or high-CHD risk groups. The values of the predictor variables for the patients with unknown status of coronary heart disease were substituted into the discriminant function and the discriminant scores were calculated. The cases were then assigned to the group whose centroid is the closest. These cases are labeled as ungrouped in

Table 4.7  Out of 50 patients 27 (54%) were classified into low-CHD risk group and 23 (46%) were classified into high-CHD group. The classification results of ungrouped cases are also presented in a histogram chart and can be seen in Appendix H.

# CHAPTER 5

## DISCUSSION AND CONCLUSIONS

### INTRODUCTION

Chapter five is the concluding chapter of this dissertation. The findings of preliminary analysis (chapter three) and further analysis (chapter four) are discussed and conclusions drawn from these discussions. The chapter also examines the limitations of the study and raises a few issues which may be the subject for further research.

### PRELIMINARY DISCUSSION

The data on 25[illegible] individuals referred [illegible] of which 2[illegible] are with known status of coronary heart disease (CHD) risk and 5[illegible] with unknown status of CHD risk were collected from the Central [illegible] Korle-Bu Teaching Hospital, Accra. The recorded data on 2[illegible] patients [illegible] belonging to the low-CHD risk group and high-CHD risk group were [illegible] to have a general [illegible] on the characteristics of the patients and [illegible] to understand the differences that exist between the two groups of patients.

The average age [illegible] 52.6 years with the highest number (22[illegible]) of cases [illegible] 59 years old. Females reported [illegible] more [illegible] (57[illegible]) [illegible] cases than males (43[illegible]) (Appendix B). However, more males [illegible] risk than females. Out of [illegible] males who reported [illegible] high-CHD risk compared to only 4[illegible]5 [illegible] females [illegible] given the general consensus that females are more cautious about their health and seek medical help more often then their male counterparts (http://www.taskforce.com).

In order to have a better picture on the differences existing between the two groups of patients, the data on the high-CHD risk group and low-CHD risk group patients were compared. The comparison of results of two groups reveal that the average age of patients in high-CHD risk group (57.51 years) is about ten years higher than in the low-CHD risk group (47.72 years). This observation is in agreement with the well-recognized fact in medicine that incidence of CHD rises steeply starting at about the age of 55 years and the disease is most common at the age of 65 years (http://www.taskforce.com).

In the study, weights and heights of the patients were collected, and based on their values, BMI was calculated by dividing weight in kilograms (kg) by height in meters squared (m²). The individuals in the low-CHD risk group were found to have normal or slightly overweight body mass, whereas the individuals in the higher-CHD risk group on the average were obese.

The four different [illegible] tests: cholesterol, triglycerides level, direct high density lipoprotein (D-HDL) and low density lipoprotein (LDL) [illegible] that are ordered together to determine risk of CHD, were also compared in the two groups of patients. We saw in chapter three (Fig 3.5 and Fig 3.6) that the concentration [illegible] in the bloodstream of patients categorized into higher-CHD risk group is higher as [illegible] low-CHD risk group patients. Whereas the average value [illegible] CHD risk group is within the normal range, the average level in the higher-CHD [illegible] normal range. The average level of triglycerides though in [illegible] the normal range, it is almost 1.5 times higher than in the low-CHD risk group. In the same way, Fig 3.7 and Fig.3.8, depict a clear difference in the values of D-HDL and LDL for the low-CHD risk and high-CHD risk groups. We observed that the LDL transports more cholesterol to the cells in high risk group patients than in the low risk group, at the same time the ability of D-HDL to remove cholesterol from the blood vessels in high-CHD risk group is less than in the low-CHD risk group. We can therefore say that higher than normal quantity of fat is stored in the body of high-CHD risk patients than in the low-CHD group.

It is a well known fact that excessive body mass is not only a risk factor for CHD but is also a cause for developing of other coronary heart disease conditions. In the study, the relationship between BMI and measured parameters of lipid profile were analysed with the purpose to see how body mass affects cholesterol level, triglycerides level and D-HDL level in the blood. It appeared (Fig 3.7 and Fig 3.8) that higher body mass index is associated with higher cholesterol level and higher triglycerides level. However as BMI increases the level of D-HDL-cholesterol ("good cholesterol"), responsible for the removal of the cholesterol from the blood vessels is decreasing. This information is a clear indication of the fact that excessive body mass affects the normal level of lipids in the blood and consequently can become a cause of hypercholesteroleamia and other risk conditions of coronary heart disease.

We also saw in Fig 3.9, that patients with normal body mass are less likely to be at risk of CHD as their CHD risk ratio is within the normal range, however, as body mass increases from overweight to severely obese, the CHD risk ratio increases above the normal range. This suggests that body mass is an important factor responsible for the development of CHD.

One of the recognized risk factors for coronary heart disease is diabetes, which is caused by high level of sugar in the blood. The influence of the level of fasting blood sugar (FBS) on CHD risk ratio was examined and we found that a sizable number of patients, belonging to the high-CHD risk group have normal levels of FBS. However, we also observed in Fig. 3.10 that patients with higher FBS levels tend to have higher CHD risk ratio.

## FURTHER DISCUSSION

As part of the study the information collected on 200 patients with known status of CHD risk was used to develop discriminant function which best separates the individuals who are at high risk of developing coronary heart disease (CHD) from individuals with low-CHD risk.

All seven variables involved in building the discriminant model: age, body mass index (BMI), serum cholesterol level, serum triglycerides level, serum direct high density lipoprotein level (D-HDL), serum low density lipoprotein (LDL) level and fasting blood sugar level (FBS) were found to be the significant predictors (Table 4.2) at significance level of 0.05, meaning that the selected discriminating variables significantly differentiate between the two groups of patients.

Since the criterion variable consists of two groups (group 1 - low-CHD risk patients and group 2 - high-CHD risk patients), only one discriminant function was estimated. The significance of discriminant function was tested and it was found to be significant beyond the 0.01 level.

The relative importance of the predictor variables was analysed from the Structure Matrix Table in Appendix G, and it appeared that cholesterol, LDL and BMI contribute most to the discriminating power of the function, with age and FBS contributing least

The validity of discriminant function was further examined The derived function was applied to classify the observations with known group membership with the purpose to see how well the discriminant function predicts group membership using the subject data from which it was derived

Furthermore, the function was successfully applied to classify the cases with unknown group membership, thus the developed discriminant model was used to classify cases with unknown status of CHD risk into low or high CHD risk groups

## CONCLUSIONS

- Majority of the patients who reported to the hospital (between August and September 2006) and who were examined for CHD risk were aged between 50 and 59 years old
- More females reported to the laboratory for their lipid profile checkup, however more males were found to be at high-CHD risk
- Patients in the high-CHD risk group on the average were ten years older than those in the low-CHD risk group
- The concentration of lipids in the bloodstream of patients categorized into high-CHD risk group is higher as compared to the low-CHD risk group patients
- Excessive body mass affects the normal level of lipids in the blood and consequently is a cause of other CHD risk conditions

- Patients with normal body mass are less likely to be at risk of CHD
- Patients with higher FBS level tend to have higher CHD risk ratio
- All seven variables involved in building the discriminant model are significant predictors for CHD risk
- Cholesterol, LDL and BMI contribute most to the discriminating power of the discriminant function, with age and FBS contributing least
- The developed discriminant function was successfully applied to classify the cases with unknown group membership into high-CHD risk or low-CHD risk groups
- The developed discriminant function can also be used to predict whether someone is likely to have a heart attack in the near future or not

## SUGGESTIONS FOR FURTHER RESEARCH

In view of the study conducted and the limitations observed, the following suggestions were found necessary for the purpose of further research and statistical accuracy

- Larger sample size and should be used in further study
- Diverse age groups should be equally covered
- The data should be collected at different time occasions with larger interval periods
- As a modified research, the data can also be collected from urban and rural areas to generate a broad picture of CHD risk conditions in the whole country, and also for comparison purposes
- The list of all known CHD risk factors should be exhausted

## REFERENCES

Lind Douglas A., Mason Robert D., Marchal William G. (2000) *Basic Statistics* 3rd ed. The McGraw-Hill Companies, Inc., USA

Zilva Joan F. and Pannall P R. (1984) *Clinical Chemistry* Lloyd-Luke (Medical Books) Ltd., London, U.K.

Pagana Kathleen Deska and Pagana Timothy J. (2003) *Diagnostic and Laboratory Test Reference* 6th ed. Elsevier Inc., Philadelphia USA

Harlow Lisa L. (2005) *The Essence of Multivariate Thinking* Lawrence Erlbaum Associates Inc., New Jersey, USA

Spiegel Murray R. (1992) *Statistics* 2nd ed. The McGraw-Hill Companies, Inc., USA

Malhotra Naresh K. (2004) *Marketing Research* 4th ed. Pearson Education, Inc., New Jersey, USA

Leech Nancy L., Barrett Karen C. Karen C. and Morgan George A. (2005) *SPSS for Intermediate Statistics* Lawrence Erlbaum Associates, Inc., New Jersey, USA

Jonson Richard A. and Wichern Dean W. (1992) *Applied Multivariate Statistical Analysis* 3rd ed. Prentice-Hall International Ltd., London, U.K.

Subhash Sharma (1996) *Applied Multivariate Techniques*. John Wiley & Sons, Inc., USA

http://www.taskforce.com/ *International Task Force for Prevention of Coronary Heart Disease*

http://www.healthguidance.org/How-to-Prevent-Coronary-Heart-Disease-and-Heart-Attack.html *How to Prevent Coronary Heart Disease and Heart Attack*

http://www.statsoft.com/texbook/stdiscan.html *Discriminant Function Analysis* http://www.ncl.ac.uk/iss/statistics/docs/discriminant.html *How to Perform and Interpret Discriminant Analysis*

http://www.spsstools.net/spss.htm-23k *SPSS Tutorial*

www.wrongdiagnosis.com/coronary-heart-disease/stats-country.htm *Statistics by country for Coronary heart disease*

# APPENDIX A

| No | Age | Gender | BMI | Choles-terol | Trigly-cerides | D-HDL | LDL | FBS | Group | Validate | Assigned to group | Discriminant score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 34 | 53 | 2 | 28.1 | 6 | 0.96 | 1.46 | 128 | 4.2 | 1 | 1 | 1 | -0.58448 |
| 35 | 53 | 2 | 31.6 | 5.78 | 1.15 | 1.41 | 151 | 5.7 | 1 | 1 | 1 | -0.15976 |
| 36 | 44 | 2 | 23.8 | 3.64 | 0.72 | 0.97 | 93 | 5.2 | 1 | 1 | 1 | -1.06622 |
| 37 | 49 | 2 | 28.3 | 5.65 | 0.98 | 1.36 | 151 | 4 | 1 | 1 | 1 | -0.43419 |
| 38 | 50 | 2 | 26.7 | 5.6 | 1.6 | 1.38 | 137 | 5.8 | 1 | 0 | 1 | -0.65694 |
| 39 | 39 | 2 | 27.9 | 5.4 | 1.1 | 1.36 | 135 | 6 | 1 | 1 | 1 | -0.64835 |
| 40 | 76 | 1 | 27.3 | 5.2 | 0.7 | 1.3 | 137 | 6.1 | 1 | 1 | 1 | -0.52843 |
| 41 | 51 | 1 | 29.1 | 5.7 | 0.74 | 1.39 | 90 | 7.1 | 1 | 1 | 1 | -0.75251 |
| 42 | 40 | 2 | 30.4 | 5.3 | 0.73 | 1.42 | 136 | 4.6 | 1 | 0 | 1 | -0.66015 |
| 43 | 54 | 2 | 25.5 | 5.7 | 1.3 | 1.4 | 146 | 4.9 | 1 | 1 | 1 | -0.73879 |
| 44 | 58 | 2 | 24.4 | 3.9 | 0.7 | 1.21 | 91 | 4.9 | 1 | 0 | 1 | -1.46726 |
| 45 | 37 | 2 | 21.9 | 3.67 | 1.37 | 1 | 78 | 6.6 | 1 | 1 | 1 | -1.3347 |
| 46 | 48 | 1 | 24.6 | 4.71 | 0.72 | 1.24 | 123 | 4.3 | 1 | 0 | 1 | -1.03707 |
| 47 | 24 | 2 | 22.2 | 4.82 | 0.82 | 1.3 | 158 | 3.7 | 1 | 1 | 1 | -1.3193 |
| 48 | 45 | 1 | 23.4 | 5.22 | 2.61 | 1.31 | 103 | 4.7 | 1 | 1 | 1 | -1.10283 |
| 49 | 57 | 1 | 22.6 | 3.07 | 0.67 | 1.4 | 156 | 4.6 | 1 | 1 | 1 | -2.35885 |
| 50 | 55 | 2 | 23.7 | 5.35 | 1.47 | 1.36 | 180 | 3.9 | 1 | 1 | 1 | -0.88199 |
| 51 | 35 | 1 | 23.3 | 5.9 | 0.54 | 1.47 | 136 | 4.5 | 1 | 1 | 1 | -1.1161 |
| 52 | 28 | 2 | 25.1 | 3.56 | 1.57 | 1.33 | 57 | 5.6 | 1 | 1 | 1 | -2.08308 |
| 53 | 22 | 2 | 23.2 | 3.57 | 1.38 | 1.15 | 69 | 5.5 | 1 | 0 | 1 | -1.70606 |
| 54 | 21 | 1 | 19.3 | 5.1 | 0.91 | 1.3 | 129 | 4 | 1 | 0 | 1 | -1.52297 |
| 55 | 22 | 2 | 23.8 | 4.3 | 1.2 | 1.39 | 94 | 5.6 | 1 | 1 | 1 | -1.86598 |
| 56 | 35 | 1 | 24.7 | 3.2 | 0.96 | 1.4 | 81 | 6.1 | 1 | 0 | 1 | -2.18553 |
| 57 | 25 | 1 | 23.2 | 3.8 | 1.2 | 1.15 | 72 | 7 | 1 | 1 | 1 | -1.59103 |
| 58 | 29 | 2 | 22.5 | 5.37 | 0.48 | 1.56 | 78 | 4.7 | 1 | 0 | 1 | -1.90568 |
| 59 | 28 | 1 | 22.2 | 6.17 | 1.44 | 1.77 | 143 | 3.8 | 1 | 1 | 1 | -1.85579 |
| 60 | 76 | 1 | 24.2 | 7.18 | 1.09 | 1.77 | 188 | 7.4 | 1 | 1 | 1 | -0.78104 |
| 61 | 25 | 1 | 22.3 | 3.22 | 0.66 | 1.36 | 60 | 6.1 | 1 | 1 | 1 | -2.58757 |
| 62 | 68 | 2 | 28.5 | 5.33 | 0.46 | 1.48 | 136 | 5.5 | 1 | 1 | 1 | -0.86425 |
| 63 | 53 | 1 | 28 | 4.36 | 0.75 | 1.38 | 155 | 5.6 | 1 | 1 | 1 | -1.11452 |
| 64 | 36 | 2 | 26.7 | 4.6 | 0.56 | 1.58 | 116 | 4.5 | 1 | 1 | 1 | -1.33391 |
| 65 | 60 | 2 | 28.5 | 5.9 | 1.19 | 1.48 | 148 | 4.6 | 1 | 1 | 1 | -0.85779 |
| 66 | 48 | 1 | 28.1 | 4.66 | 0.89 | 1.6 | 103 | 3.9 | 1 | 0 | 1 | -1.76616 |
| 67 | 61 | 2 | 25.6 | 3.65 | 0.85 | 1.48 | 144 | 4.6 | 1 | 1 | 1 | -1.01967 |
| 68 | 51 | 2 | 21.9 | 3.72 | 1.34 | 1.36 | 67 | 6.6 | 1 | 0 | 1 | [illegible] |
| 69 | 35 | 2 | 19.6 | 4.56 | 0.65 | 1.38 | 110 | 4.8 | 1 | 1 | 1 | -1.94598 |
| 70 | 68 | 1 | 22.5 | 5.82 | 1.44 | 1.46 | 112 | 6 | 1 | 1 | 1 | 1.012 |

| No | Age | Gender | BMI | Choles-terol | Trigly-cerides | D-HDL | LDL | FBS | Group | Validate | Assigned to group | Discriminant score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 71 | 41 | 1 | 23.2 | 4.32 | 0.78 | 1.41 | 97 | 5.9 | 1 | 1 | 1 | -1.88637 |
| 72 | 51 | 2 | 22.9 | 7.02 | 1.59 | 1.95 | 24 | 4.8 | 1 | 0 | 1 | -2.10157 |
| 73 | 31 | 1 | 25.2 | 6.6 | 0.98 | 1.64 | 173 | 3.7 | 1 | 1 | 1 | -0.92314 |
| 74 | 59 | 2 | 30.5 | 6.37 | 1.16 | 2.03 | 146 | 4.9 | 1 | 0 | 1 | -1.56799 |
| 75 | 40 | 1 | 20.6 | 5.18 | 0.9 | 1.74 | 115 | 5.8 | 1 | 1 | 1 | -2.46298 |
| 76 | 53 | 2 | 23.4 | 6.44 | 1.17 | 1.74 | 159 | 6.4 | 1 | 1 | 1 | -1.35071 |
| 77 | 39 | 1 | 21.9 | 6.19 | 1.12 | 1.87 | 145 | 6.1 | 1 | 0 | 1 | -2.04952 |
| 78 | 47 | 2 | 25.3 | 5.53 | 1.1 | 1.69 | 128 | 3.8 | 1 | 1 | 1 | -1.70362 |
| 79 | 51 | 2 | 30.1 | 6.17 | 1.17 | 1.62 | 153 | 4.1 | 1 | 1 | 1 | -0.6599 |
| 80 | 48 | 2 | 28.3 | 3.09 | 1.14 | 1.21 | 86 | 4.6 | 1 | 1 | 1 | -1.12576 |
| 81 | 29 | 1 | 23.3 | 5.94 | 1.12 | 1.87 | 135 | 5.1 | 1 | 1 | 1 | -2.14148 |
| 82 | 46 | 2 | 28.9 | 6.76 | 1.13 | 1.72 | 173 | 6.2 | 1 | 1 | 1 | -0.62619 |
| 83 | 41 | 1 | 28.4 | 4.46 | 0.7 | 1.46 | 103 | 3.9 | 1 | 0 | 1 | -1.50239 |
| 84 | 35 | 2 | 26.3 | 4.66 | 1.06 | 1.51 | 101 | 4.6 | 1 | 1 | 1 | -1.72055 |
| 85 | 26 | 2 | 22.9 | 3.7 | 0.99 | 1.1 | 82 | 4.5 | 1 | 1 | 1 | -1.555 |
| 86 | 58 | 1 | 25 | 5.88 | 0.69 | 1.51 | 155 | 5.7 | 1 | 1 | 1 | -0.92299 |
| 87 | 31 | 1 | 28.4 | 4.96 | 1.19 | 1.33 | 118 | 5.4 | 1 | 0 | 1 | -0.85035 |
| 88 | 77 | 2 | 26.4 | 5.38 | 0.69 | 1.49 | 137 | 5.1 | 1 | 1 | 1 | -1.02146 |
| 89 | 97 | 2 | 24.3 | 4.58 | 2.21 | 1.21 | 90 | 4.5 | 1 | 1 | 1 | -0.98636 |
| 90 | 58 | 2 | 25.2 | 5.9 | 0.36 | 1.72 | 154 | 3.9 | 1 | 1 | 1 | -1.48024 |
| 91 | 45 | 2 | 27.9 | 6.16 | 2.28 | 1.74 | 129 | 6 | 1 | 0 | 1 | -1.22253 |
| 92 | 26 | 2 | 20.8 | 2.81 | 0.78 | 0.77 | 64 | 3.9 | 1 | 1 | 1 | -1.42451 |
| 93 | 30 | 1 | 24 | 4.34 | 1.48 | 1.1 | 98 | 4.2 | 1 | 0 | 1 | -1.05478 |
| 94 | 29 | 1 | 25 | 4.55 | 1.16 | 1.38 | 101 | 4.6 | 1 | 1 | 1 | -1.57325 |
| 95 | 23 | 2 | 22.5 | 4.47 | 0.8 | 1.15 | 114 | 5.1 | 1 | 1 | 1 | -1.20875 |
| 96 | 22 | 1 | 25.8 | 5.07 | 0.92 | 1.49 | 121 | 5.3 | 1 | 0 | 1 | -1.46049 |
| 97 | 59 | 2 | 33.2 | 4.91 | 0.48 | 1.23 | 133 | 6.1 | 1 | 1 | 1 | -0.05063 |
| 98 | 25 | 2 | 19.5 | 4.13 | 1.57 | 1.28 | 82 | 5.3 | 1 | 0 | 1 | -2.07779 |
| 99 | 42 | 2 | 28.9 | 3.95 | 0.83 | 1.13 | 120 | 8.1 | 1 | 0 | 1 | -0.73479 |
| 100 | 57 | 1 | 29.1 | 4.6 | 1.04 | 1.28 | 109 | 7.2 | 1 | 1 | 1 | -0.76451 |
| 101 | 49 | 1 | 30.1 | 6.09 | 1.67 | 1.06 | 93 | 4.6 | 2 | 1 | 2 | 0.568 |
| 102 | 57 | 1 | 36.9 | 6.24 | 2.6 | 1.19 | 89 | 5.6 | 2 | 1 | 2 | 0.94733 |
| 103 | 58 | 2 | 39 | 6.77 | 1.69 | 0.96 | 101 | 6.5 | 2 | 1 | 2 | 2.06113 |
| 104 | 82 | 2 | 27.2 | 8.18 | 3 | 0.89 | 121 | 4.2 | 2 | 0 | 2 | 2.06771 |
| 105 | 46 | 2 | 35.3 | 7.3 | 2.39 | 1.23 | 191 | 5.7 | 2 | 0 | 2 | 1.55961 |
| 106 | 51 | 1 | 35.9 | 6.35 | 1.48 | 1.05 | 177 | 3.7 | 2 | 1 | 2 | 1.50224 |
| 107 | 62 | 2 | 36.2 | 7.16 | 2.53 | 1.28 | 178 | 6.3 | 2 | 1 | 2 | 1.45574 |

| No | Age | Gender | BMI | [illegible] | [illegible] | [illegible] | [illegible] | FBS | Group | [illegible] | [illegible] | [illegible] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 108 | 59 | 2 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | 62 | 2 | [illegible] | [illegible] | | [illegible] | [illegible] | [illegible] | | [illegible] | | [illegible] |
| [illegible] | 59 | 2 | 34.2 | 8 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 2 | [illegible] |
| [illegible] | 47 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | | | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | | | [illegible] |
| [illegible] | 68 | 1 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | [illegible] | | [illegible] | | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | | [illegible] | | | | [illegible] | [illegible] | | | | [illegible] |
| [illegible] | [illegible] | | [illegible] | | | [illegible] | [illegible] | | | | | [illegible] |
| | [illegible] | | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | | | | |
| [illegible] | [illegible] | | [illegible] | [illegible] | | [illegible] | [illegible] | [illegible] | | | [illegible] | [illegible] |
| [illegible] | [illegible] | | [illegible] | [illegible] | | [illegible] | [illegible] | | | | | [illegible] |
| [illegible] | [illegible] | | [illegible] | | | | [illegible] | [illegible] | | | | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | | [illegible] | [illegible] | | | | [illegible] |
| | [illegible] | | [illegible] | [illegible] | | [illegible] | [illegible] | | | | | [illegible] |
| | [illegible] | | [illegible] | [illegible] | | | [illegible] | | | | | [illegible] |
| [illegible] | [illegible] | | | [illegible] | [illegible] | | | [illegible] | | | | [illegible] |
| | [illegible] | | | | [illegible] | | [illegible] | [illegible] | | | | |
| [illegible] | [illegible] | | | | | | | | | | | [illegible] |
| [illegible] | [illegible] | | | [illegible] | | | | [illegible] | | | | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | | | [illegible] | | | | | [illegible] |
| [illegible] | [illegible] | | | | [illegible] | [illegible] | [illegible] | [illegible] | | | | |
| [illegible] | [illegible] | | [illegible] | [illegible] | | | | | | | | [illegible] |
| [illegible] | [illegible] | | | | | [illegible] | [illegible] | [illegible] | | | | |
| [illegible] | [illegible] | | [illegible] | [illegible] | | | | | | | | |
| [illegible] | [illegible] | | [illegible] | | | | | | | | | |
| [illegible] | | | [illegible] | [illegible] | | | [illegible] | | | | | |
| [illegible] | [illegible] | [illegible] | [illegible] | | [illegible] | | | | | | | |
| [illegible] | [illegible] | | | [illegible] | | | | | | | | |
| [illegible] | [illegible] | | [illegible] | [illegible] | | | [illegible] | | | | | |
| [illegible] | | | | | | | | [illegible] | | | | |
| [illegible] | | | | | | | [illegible] | | | | | |
| [illegible] | [illegible] | | | [illegible] | | [illegible] | | [illegible] | | | | |
| [illegible] | [illegible] | | [illegible] | | | | | | | | | |
| [illegible] | [illegible] | [illegible] | | | | | [illegible] | | | | | |
| [illegible] | 68 | | [illegible] | | [illegible] | [illegible] | | [illegible] | | | | |
| 144 | [illegible] | [illegible] | [illegible] | | [illegible] | | | | | | | |

| No | Age | Gender | BMI | Cholesterol | Triglycerides | D-HDL | LDL | FBS | Group | Validate | Assigned to group | Discriminant score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 145 | 77 | 1 | 28 | 5.74 | 1.27 | 1.29 | 148 | 5.3 | 2 | 0 | 1 | -0.1245 |
| 146 | 51 | 1 | 26.1 | 5.3 | 0.4 | 0.98 | 159 | 4.4 | 2 | 0 | 2 | 0.20915 |
| 147 | 70 | 2 | 38.5 | 7.1 | 1.5 | 0.72 | 228 | 4.8 | 2 | 1 | 2 | 3.23338 |
| 148 | 60 | 2 | 33.3 | 6.8 | 1.2 | 1.37 | 188 | 6 | 2 | 1 | 2 | 0.78306 |
| 149 | 87 | 2 | 37.9 | 6.87 | 0.96 | 0.91 | 161 | 4.7 | 2 | 0 | 2 | 2.39468 |
| 150 | 36 | 1 | 34.7 | 5.9 | 0.98 | 1.36 | 196 | 5.2 | 2 | 1 | 2 | 0.38724 |
| 151 | 45 | 2 | 32.4 | 6.7 | 0.8 | 1.21 | 197 | 4.3 | 2 | 1 | 2 | 1.01357 |
| 152 | 62 | 2 | 28.8 | 7.68 | 0.67 | 1.79 | 213 | 6 | 2 | 1 | 1 | -0.12649 |
| 153 | 79 | 1 | 32 | 7.1 | 2 | 1.21 | 191 | 5.9 | 2 | 0 | 2 | 1.32059 |
| 154 | 56 | 2 | 32.1 | 5.9 | 1 | 1.38 | 156 | 5.1 | 2 | 1 | 2 | 0.03937 |
| 155 | 61 | 1 | 30.2 | 4.7 | 1.37 | 0.9 | 121 | 4.6 | 2 | 1 | 2 | 0.38168 |
| 156 | 53 | 2 | 41 | 7.6 | 0.9 | 1.07 | 234 | 7.3 | 2 | 1 | 2 | 2.81268 |
| 157 | 68 | 2 | 28.7 | 6.4 | 1.1 | 1.31 | 176 | 5.2 | 2 | 0 | 2 | 0.29358 |
| 158 | 67 | 2 | 29.7 | 8.35 | 1.79 | 1.21 | 243 | 5 | 2 | 0 | 2 | 1.89303 |
| 159 | 85 | 2 | 33.9 | 9.45 | 1.5 | 1.08 | 347 | 6.1 | 2 | 1 | 2 | 3.59556 |
| 160 | 83 | 2 | 36 | 7.6 | 1.42 | 2.03 | 267 | 6 | 2 | 1 | 2 | 2.66096 |
| 161 | 56 | 1 | 27.8 | 6.27 | 1.37 | 1.33 | 217 | 5.6 | 2 | 1 | 2 | 0.19761 |
| 162 | 52 | 1 | 36 | 7.55 | 1.2 | 1.15 | 269 | 7 | 2 | 1 | 2 | 2.24556 |
| 163 | 56 | 1 | 20.5 | 4.69 | 0.57 | 1.05 | 129 | 4.9 | 2 | 1 | 1 | -0.85059 |
| 164 | 60 | 1 | 27.7 | 6.25 | 1.11 | 1.38 | 166 | 6 | 2 | 1 | 1 | [illegible] |
| 165 | 56 | 2 | 29.7 | 5.97 | 0.75 | 1.21 | 170 |  | 2 | 0 | 2 | 0.32157 |
| 166 | 53 | 1 | 28 | 6.77 | 0.9 | 1.38 | 190 | 6.2 | 2 | 1 | 2 | 0.25973 |
| 167 | 58 | 1 | 28.7 | 10.41 | 3 | 1.67 | 279 | 5.9 | 2 | 1 | 2 | 1.81702 |
| 168 | 72 | 1 | 25.8 | 5.17 | 0.54 | 1.5 | 130 | 5.3 | 2 | 0 | 1 | -1.24815 |
| 169 | 56 | 1 | 27.3 | 7.18 | 2.6 | 1.12 | 186 | 4.3 | 2 | 0 | 2 | 1.07089 |
| 170 | 51 | 2 | 33.9 | 6.45 | 2.72 | 1.26 | 150 | 4.4 | 2 | 1 | 2 | 0.77038 |
| 171 | 58 | 2 | 32.3 | 8.7 | 1.36 | 1.19 | 230 | 5.9 | 2 | 1 | 2 | 2.30073 |
| 172 | 55 | 1 | 29.4 | 6.5 | 0.67 | 1.26 | 189 | 4.1 | 2 | 0 | 2 | 0.51467 |
| 173 | 70 | 2 | 31.6 | 6.5 | 0.43 | 1.3 | 163 | 7.2 | 2 | 0 | 2 | 0.59578 |
| 174 | 58 | 2 | 21.6 | 4.11 | 0.81 | 0.92 | 108 | 7 | 2 | 0 | 1 | -0.75748 |
| 175 | 68 | 1 | 25.5 | 8.48 | 1.11 | 1.46 | 250 | 5.1 | 2 | 1 | 2 | 0.96317 |
| 176 | 71 | 2 | 28.8 | 6.03 | 2.34 | 1.05 | 150 | 4.6 | 2 | 1 | 2 | 0.70938 |
| 177 | 36 | 1 | 26 | 6.69 | 0.69 | 1.54 | 185 | 5.3 | 2 | 1 | 1 | -0.46384 |
| 178 | 29 | 1 | 30.5 | 7.13 | 1.04 | 1.64 | 188 | 4.5 | 2 | [illegible] | 1 | -0.21851 |
| 179 | 74 | 1 | 26.5 | 6.56 | 1.2 | 1.51 | 172 | 5.8 | 2 | 0 | 1 | -0.32772 |
| 180 | 31 | 1 | 30.4 | 6.84 | 0.62 | 1.44 | 196 | 6 | 2 | 1 | 2 | 0.29458 |
| 181 | 52 | 1 | 30.8 | 6.92 | 2.15 | 0.74 | 199 | 4.6 | 2 | 1 | 2 | [illegible] |

| No | Age | Gender | BMI | Choles-<br>terol | Trigly-<br>cerides | HDL | LDL | FBS | Group | Validate | Assigned<br>to group | Discriminant<br>score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 182 | 68 | 1 | 33.9 | 7.33 | 1.78 | 1.23 | 203 | 6.1 | 2 | 0 | 2 | 1.56286 |
| 183 | 59 | 2 | 33.2 | 6.94 | 1.53 | 1.41 | 185 | 5.7 | 2 | 1 | 2 | 0.72934 |
| 184 | 48 | 1 | 27.1 | 5.43 | 3 | 1.15 | 109 | 5.3 | 2 | 0 | 1 | -0.20598 |
| 185 | 61 | 2 | 27.2 | 6.68 | 1.21 | 1.08 | 194 | 4.2 | 2 | 1 | 2 | 0.92741 |
| 186 | 38 | 2 | 35.4 | 11.1 | 2.44 | 0.87 | 350 | 5.7 | 2 | 1 | 2 | 5.01061 |
| 187 | 57 | 2 | 22.1 | 7.67 | 1.3 | 1.41 | 217 | 5.8 | 2 | 1 | 2 | 0.6768 |
| 188 | 46 | 1 | 27.1 | 5.2 | 1.46 | 0.64 | 149 | 5.4 | 2 | 0 | 2 | 1.10876 |
| 189 | 65 | 2 | 30.1 | 6.68 | 1.57 | 0.79 | 198 | 6.4 | 2 | 1 | 2 | 2.06614 |
| 190 | 63 | 2 | 33.6 | 5.77 | 0.62 | 0.92 | 175 | 6.8 | 2 | 1 | 2 | 1.41015 |
| 191 | 26 | 2 | 28.3 | 6.27 | 0.9 | 1.44 | 169 | 5.8 | 2 | 0 | 1 | -0.29683 |
| 192 | 71 | 1 | 27.7 | 6.66 | 3 | 1.03 | 156 | 4.9 | 2 | 1 | 2 | 1.02999 |
| 193 | 47 | 2 | 27.2 | 6.94 | 1.24 | 1 | 206 | 7.8 | 2 | 1 | 2 | 1.33312 |
| 194 | 58 | 1 | 28.4 | 5.69 | 1 | 0.85 | 168 | 4 | 2 | 1 | 2 | 1.00672 |
| 195 | 45 | 2 | 25.9 | 4.6 | 1.84 | 0.92 | 109 | 4.9 | 2 | 0 | 1 | -0.18163 |
| 196 | 62 | 1 | 29.4 | 8.35 | 3 | 1.36 | 254 | 4.5 | 2 | 0 | 2 | 1.49871 |
| 197 | 42 | 2 | 30.5 | 6.24 | 1.46 | 1.13 | 170 | 6 | 2 | 0 | 2 | 0.2508 |
| 198 | 58 | 2 | 26.8 | 6.55 | 1.28 | 1.08 | 188 | 5.8 | 2 | 0 | 2 | 0.82393 |
| 199 | 55 | 2 | 33.8 | 6.71 | 1.14 | 1.18 | 192 | 4.8 | 2 | 1 | 2 | 1.24825 |
| 200 | 57 | 1 | 28.7 | 4.03 | 1.14 | 0.9 | 100 | 5.1 | 2 | 0 | 1 | -0.17966 |
| 201 | 70 | 2 | 30.9 | 5.49 | 1.14 | 1.38 | 137 | 4.6 |  |  | 1 | -0.30813 |
| 202 | 31 | 2 | 25.7 | 4.16 | 1.14 | 1.38 | 86 | 4.5 |  |  | 1 | -1.75731 |
| 203 | 65 | 2 | 30.4 | 6.37 | 1.14 | 1.54 | 165 | 5.4 |  |  | 1 | -0.21238 |
| 204 | 22 | 2 | 21.4 | 3.51 | 0.43 | 1.1 | 84 | 4.4 |  |  | 1 | -1.80677 |
| 205 | 74 | 2 | 32.4 | 6.42 | 1.7 | 1.46 | 160 | 3.9 |  |  | 2 | 0.18967 |
| 206 | 30 | 1 | 25.2 | 5.53 | 1.06 | 2.05 | 103 | 1.6 |  |  | 1 | -2.77551 |
| 207 | 33 | 1 | 29.4 | 4.78 | 0.64 | 1.37 | 139 | 5.1 |  |  | 1 | -0.89839 |
| 208 | 50 | 2 | 31.6 | 5.93 | 1.41 | 1.26 | 169 | 6 |  |  | 2 | 0.36672 |
| 209 | 58 | 1 | 31.5 | 7.49 | 1.01 | 1.38 | 216 | 5.8 |  |  | 2 | 1.04553 |
| 210 | 70 | 1 | 26.6 | 8.32 | 1.35 | 1.54 | 235 | 5 |  |  | 2 | 0.72525 |
| 211 | 52 | 1 | 28.4 | 3.87 | 0.62 | 0.92 | 132 | 4.7 |  |  | 1 | -1.26771 |
| 212 | 26 | 1 | 31.1 | 6.4 | 0.84 | 0.97 | 193 | 5.7 |  |  | 2 | 1.31655 |
| 213 | 62 | 2 | 33.2 | 7.2 | 2.27 | 1.36 | 184 | 5.9 |  |  | 2 | 1.01456 |
| 214 | 38 | 2 | 26.4 | 2.5 | 1.01 | 0.95 | 55 | 4.1 |  |  | 1 | [illegible] |
| 215 | 65 | 2 | 28.4 | 5.08 | 0.83 | 1.31 | 164 | 5.9 |  |  | 1 | [illegible] |
| 216 | 39 | 1 | 35 | 8.37 | 1.57 | 1.08 | 252 | 6 |  |  | 2 | [illegible] |
| 217 | 56 | 2 | 30.8 | 5.38 | 1.95 | 1.46 | 116 | 6.2 |  |  | 2 | 0.65926 |
| 218 | 66 | 2 | 28.2 | 5.59 | 1.28 | 1.59 | 131 | 4.7 |  |  | 2 | 1.06798 |

| No | Age | Gender | BMI | Choles-terol | Trigly-cerides | D-HDL | LDL | FBS | Group | Validate | Assigned to group | Discriminant score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 219 | 75 | 1 | 29 | 6.45 | 1.11 | 1.1 | 186 | 5 | | | 2 | 0.94408 |
| 220 | 70 | 1 | 32.6 | 7.15 | 3.15 | 0.41 | 193 | 4 | | | 2 | 3.43577 |
| 221 | 75 | 2 | 32 | 4.94 | 1.19 | 0.87 | 135 | 4.3 | | | 2 | 0.82829 |
| 222 | 38 | 2 | 24 | 3.51 | 1.03 | 1.26 | 98 | 3.9 | | | 1 | -1.89566 |
| 223 | 49 | 2 | 33.9 | 6.47 | 1.06 | 0.97 | 192 | 4.6 | | | 2 | 1.65249 |
| 224 | 69 | 2 | 32.4 | 7.33 | 1.63 | 1.67 | 188 | 6.2 | | | 2 | 0.24115 |
| 225 | 60 | 2 | 32.3 | 5.51 | 1.13 | 1.1 | 149 | 7.8 | | | 2 | 0.61889 |
| 226 | 30 | 2 | 21.2 | 4.21 | 0.57 | 1.49 | 94 | 4.9 | | | 1 | -2.39386 |
| 227 | 62 | 1 | 29 | 7.15 | 1.05 | 1.67 | 192 | 5.5 | | | 1 | -0.18121 |
| 228 | 78 | 1 | 25.5 | 4.26 | 1.05 | 1.21 | 99 | 6 | | | 1 | -1.06632 |
| 229 | 50 | 1 | 30.1 | 4.97 | 0.8 | 0.95 | 115 | 6.1 | | | 2 | 0.35309 |
| 230 | 56 | 2 | 35.8 | 9.33 | 1.82 | 1.31 | 276 | 4.8 | | | 2 | 2.75766 |
| 231 | 61 | 1 | 26.4 | 5.75 | 0.81 | 1.05 | 166 | 5 | | | 2 | 0.36167 |
| 232 | 65 | 1 | 27.7 | 5.15 | 0.92 | 1.44 | 126 | 6.4 | | | 1 | -0.94721 |
| 233 | 56 | 2 | 33.4 | 6.42 | 3.24 | 0.95 | 153 | 8 | | | 2 | 1.61305 |
| 234 | 70 | 1 | 22.1 | 4.95 | 0.85 | 1.69 | 109 | 5.1 | | | 1 | -2.25964 |
| 235 | 28 | 2 | 21 | 4.71 | 0.7 | 1.95 | 93 | 4.9 | | | 1 | -3.35135 |
| 236 | 63 | 1 | 27.7 | 5.48 | 1.88 | 1.46 | 121 | 3.9 | | | 1 | -0.88271 |
| 237 | 31 | 2 | 25.8 | 5.17 | 1.64 | 1.64 | 106 | 4 | | | 1 | -1.83402 |
| 238 | 25 | 1 | 23.5 | 4.99 | 1.21 | 1.26 | 122 | 5.2 | | | 1 | -1.093 |
| 239 | 44 | 2 | 28.4 | 4.21 | 0.52 | 1.44 | 97 | 6.1 | | | 1 | -1.55469 |
| 240 | 75 | 1 | 28.1 | 6.37 | 1.81 | 0.97 | 175 | 7.8 | | | 2 | 1.18268 |
| 241 | 56 | 2 | 27.9 | 7.44 | 1.88 | 1.79 | 183 | 5 | | | 1 | -0.18561 |
| 242 | 60 | 2 | 31.5 | 6.14 | 0.83 | 1.36 | 168 | 11 | | | 2 | 0.32288 |
| 243 | 61 | 1 | 29.1 | 6.08 | 1.63 | 1.36 | 152 | 4.1 | | | 1 | -0.08593 |
| 244 | 54 | 1 | 28.7 | 3.48 | 0.66 | 0.64 | 97 | 4.8 | | | 2 | 0.17558 |
| 245 | 24 | 1 | 24 | 5.64 | 0.79 | 1.64 | 139 | 5.3 | | | 1 | -1.64329 |
| 246 | 25 | 2 | 31.2 | 6.89 | 2.75 | 0.69 | 155 | 5.9 | | | 2 | 2.20986 |
| 247 | 37 | 1 | 26.8 | 4.82 | 1.58 | 1.38 | 103 | 6.1 | | | 1 | -1.30833 |
| 248 | 48 | 1 | 34.5 | 7.29 | 1.05 | 1.33 | 210 | 5.2 | | | 2 | 1.26952 |
| 249 | 46 | 2 | 32.6 | 5.82 | 1.12 | 1.23 | 156 | 4 | | | 2 | 0.38043 |
| 250 | 31 | 2 | 24.7 | 5.11 | 0.9 | 1.31 | 130 | 4.8 | | | 1 | -1.01906 |

**The units of measurements:**

BMI - $\frac{kg}{m^2}$

Cholesterol - mmol/L.

Triglycerides - mmol/L.

D-HDL - mmol/L.

LDL – mg/dL

FBS - mmol/L.

## APPENDIX B

### B1: Some recommendations that help to lower the risk of coronary heart disease (http://www.healthguidance.org/How-to-Prevent-Coronary-Heart-Disease-and-Heart-Attack.html/)

i. **Quit smoking** Cigarette smoking is considered the biggest risk factor for sudden cardiac death A smokers' risk of heart attack is more than twice that of a nonsmoker

ii **Control the blood pressure** High blood pressure causes the heart to work harder Over time, the heart enlarges This increases the risk of heart attack and other related disease

conditions such as congestive heart failure, stroke and kidney failure

iii **Control the blood cholesterol** As levels of choles'erol rise in the blood, the risk for CHD increases By adopting a diet that is lower in saturated fat and cholesterol, it is possible to reduce the level of cholesterol

iv **Increase physical activity** Moderate exercise, when done regularly, can reduce the risk for heart disease Regular exercise can also help to reduce the cholesterol levels and blood pressure, as well as decrease the risk for developing diabetes and obesity

v **Maintain a desirable body weight.** People with excess body fat are more likely to develop CHD. Obesity not only puts too much strain on the heart but it can also affect blood pressure and cholesterol and increase the risk for diabetes

vi. **Maintain normal blood sugar levels.** Diabetes increases the risk of developing heart disease, even when the blood sugars are kept under control

vii **Reduce stress.** Some researchers have noted a connection between CHD and stress

Although scientists still do not know exactly how stress might increase the risk for heart disease, it is generally considered wise to try to avoid stress as much as possible

## B2: Precautionary measures taken for reliable laboratory results.

The following precautionary measures were taken or observed by qualified staff so as to generate reliable and accurate laboratory results

**Reagent supply**

Up to date reagent packs were purchased from well known and reliable manufactures. These reagents were properly assessed and stored according to the manufacturer's specifications

**Maintenance of equipments**

Daily and periodic maintenance of the laboratory equipment were carried out. Calibration and standardization of chemistry analyzers ensure accurate results for patients. Control samples of known concentrations were also measured before and alongside samples of patients

Specifically, in estimating the blood lipid profile of patients, the following measures were taken in accordance with international standards to ensure reliable results

i) Patients were advised to fast overnight (about 8 -10 hours) before blood samples were drawn

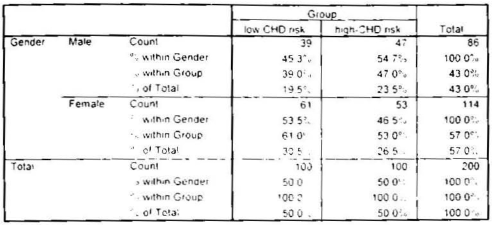
ii) The blood samples were treated with care to avoid the break-up of the red blood cells into the serum which could alter the actual concentrations of the lipid profiles in the serum

iii) Samples that were not worked on the same day, i.e. day the patient's sample was taken, were stored under the required laboratory procedures

iv) Results obtained were carefully recorded by competent staff

# APPENDIX C

Gender * Group Crosstabulation



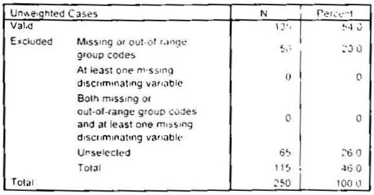
| | | | Group | | Total |
|---|---|---|---|---|---|
| | | | low CHD risk | high-CHD risk | |
| Gender | Male | Count | 39 | 47 | 86 |
| | | % within Gender | 45.3% | 54.7% | 100.0% |
| | | % within Group | 39.0% | 47.0% | 43.0% |
| | | % of Total | 19.5% | 23.5% | 43.0% |
| | Female | Count | 61 | 53 | 114 |
| | | % within Gender | 53.5% | 46.5% | 100.0% |
| | | % within Group | 61.0% | 53.0% | 57.0% |
| | | % of Total | 30.5% | 26.5% | 57.0% |
| Total | | Count | 100 | 100 | 200 |
| | | % within Gender | 50.0% | 50.0% | 100.0% |
| | | % within Group | 100.0% | 100.0% | 100.0% |
| | | % of Total | 50.0% | 50.0% | 100.0% |

## APPENDIX D

### The process of division of the cases with known status of CHD risk

The sample examined in this study work consists of 250 patients laboratory results, collected from the Central Laboratory of Korle-Bu Teaching Hospital, Accra. Out of the total observed cases - 200 were patients with known coronary heart disease risk and 50 with unknown status of CHD risk. The 200 cases with known status of CHD risk were randomly divided such that 70% of the known risk cases were used for the estimation of the discriminant function and the other 30% of the known risk cases were reserved to validate the discriminant function. The random division was done by creating a new variable "validate". The set of the values of "validate" was randomly generated by Bernoulli variates with probability parameter 0.7. The Bernoulli variate takes value of 1 with probability 0.7 and value 0 with probability 0.3. As we can see from the data table in Appendix A, column 11, 70% (135 cases with known status of CHD) have a "validate" value of 1 and 30% (65 cases with known status of CHD) have a "validate" value of 0. Furthermore, the 135 known risk cases were used as analysis sample for the estimation of the discriminant function. The other 65 known risk cases served as holdout sample and were reserved to validate the discriminant function.

Analysis Case Processing Summary



| Unweighted Cases | | N | Percent |
|---|---|---|---|
| Valid | | 135 | 54.0 |
| Excluded | Missing or out-of range group codes | 50 | 20.0 |
| | At least one missing discriminating variable | 0 | 0 |
| | Both missing or out-of-range group codes and at least one missing discriminating variable | 0 | 0 |
| | Unselected | 65 | 26.0 |
| | Total | 115 | 46.0 |
| Total | | 250 | 100.0 |

## APPENDIX E

### E1: Box's Test of Equality of Covariance Matrices

**Log Determinants**

| Group | Rank | Log Determinant |
|---|---|---|
| Low-CHD risk | 7 | 8 633 |
| High-CHD risk | 7 | 11 019 |
| Pooled within-groups | 7 | 10 465 |

The ranks and natural logarithms of determinants printed are those of the group covariance matrices

**Test Results**

| | | |
|---|---|---|
| Box s M | | 86 163 |
| F | Approx | 2 905 |
| | df1 | 28 |
| | df2 | 61609 912 |
| | Sig | 000 |

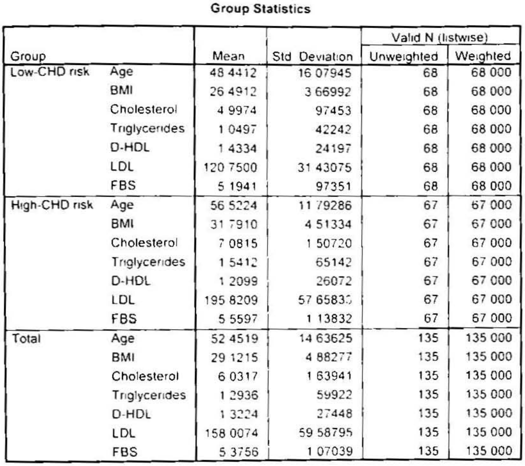
Tests null hypothesis of equal population covariance matrices

### E2: Pearson Correlations

| | | Age | BMI | Cholesterol | Triglyceride | D-HDL | LDL | FBS |
|---|---|---|---|---|---|---|---|---|
| Age | | 1 | 365** | 313** | 190** | 081 | 294** | 060 |
| | Sig (2-tailed) | | 000 | 000 | 007 | 256 | 000 | 401 |
| BMI | | 365** | 1 | 509** | 307** | - 126 | 478** | 119 |
| | Sig (2-tailed) | 000 | | 000 | 000 | 075 | 000 | 092 |
| Cholesterol | | 313** | 509** | 1 | 498** | 025 | 886** | 117 |
| | Sig (2-tailed) | 000 | 000 | | 000 | 723 | 000 | 098 |
| Triglycerides | | 190** | 307** | 498** | 1 | - 192** | 323** | 065 |
| | Sig (2-tailed) | 007 | 000 | 000 | | 006 | 000 | 361 |
| D-HDL | | - 081 | - 126 | 025 | - 192** | 1 | - 112 | - 064 |
| | Sig (2-tailed) | 256 | 075 | 723 | 006 | | 113 | 366 |
| LDL | | 294** | 478** | 886** | 323** | 112 | 1 | 110 |
| | Sig (2-tailed) | 000 | 000 | 000 | 000 | 113 | | 121 |
| FBS | | 060 | 119 | 117 | 065 | 064 | 110 | 1 |
| | Sig (2-tailed) | 401 | 092 | 098 | 361 | 366 | 121 | |

** Correlation is significant at the 0 01 level (2-tailed)

## APPENDIX F



**Group Statistics**

| Group | | Mean | Std Deviation | Valid N (listwise) | |
|---|---|---|---|---|---|
| | | | | Unweighted | Weighted |
| Low-CHD risk | Age | 48 4412 | 16 07945 | 68 | 68 000 |
| | BMI | 26 4912 | 3 66992 | 68 | 68 000 |
| | Cholesterol | 4 9974 | 97453 | 68 | 68 000 |
| | Triglycerides | 1 0497 | 42242 | 68 | 68 000 |
| | D-HDL | 1 4334 | 24197 | 68 | 68 000 |
| | LDL | 120 7500 | 31 43075 | 68 | 68 000 |
| | FBS | 5 1941 | 97351 | 68 | 68 000 |
| High-CHD risk | Age | 56 5224 | 11 79286 | 67 | 67 000 |
| | BMI | 31 7910 | 4 51334 | 67 | 67 000 |
| | Cholesterol | 7 0815 | 1 50720 | 67 | 67 000 |
| | Triglycerides | 1 5412 | 65142 | 67 | 67 000 |
| | D-HDL | 1 2099 | 26072 | 67 | 67 000 |
| | LDL | 195 8209 | 57 65833 | 67 | 67 000 |
| | FBS | 5 5597 | 1 13832 | 67 | 67 000 |
| Total | Age | 52 4519 | 14 63625 | 135 | 135 000 |
| | BMI | 29 1215 | 4 88277 | 135 | 135 000 |
| | Cholesterol | 6 0317 | 1 63941 | 135 | 135 000 |
| | Triglycerides | 1 2936 | 59922 | 135 | 135 000 |
| | D-HDL | 1 3224 | 27448 | 135 | 135 000 |
| | LDL | 158 0074 | 59 58795 | 135 | 135 000 |
| | FBS | 5 3756 | 1 07039 | 135 | 135 000 |

## APPENDIX G

**Standardized Canonical Discriminant Function Coefficients**

| | Function |
|---|---|
| | 1 |
| Age | 045 |
| BMI | 362 |
| Cholesterol | 672 |
| Triglycerides | 005 |
| D-HDL | - 654 |
| LDL | 150 |
| FBS | 019 |

**Structure Matrix**

| | Function |
|---|---|
| | 1 |
| Cholesterol | 676 |
| LDL | 666 |
| BMI | 530 |
| Triglycerides | 368 |
| D-HDL | - 365 |
| Age | 235 |
| FBS | 142 |

**Canonical Discriminant Function Coefficients**

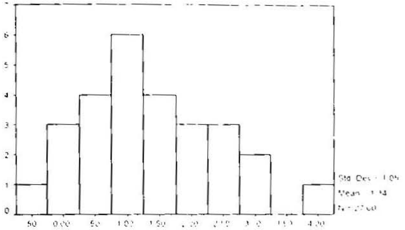
| | Function |
|---|---|
| | 1 |
| Age | 003 |
| BMI | 088 |
| Cholesterol | 531 |
| Triglycerides | 010 |
| D-HDL | -2 602 |
| LDL | 003 |
| FBS | 018 |
| (Constant) | -3 110 |

Unstandardized coefficients

## APPENDIX II

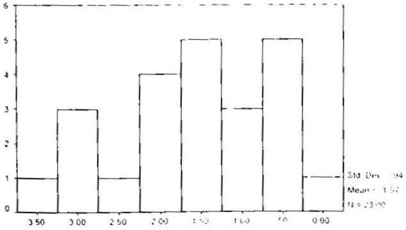
Ungrouped cases

Classified into low-CHD risk group



Ungrouped cases

Classified into high-CHD risk group


Std. Dev = .94
Mean = 1.57
N = 23.00