Georgia State University
Digital Archive @ GSU

Mathematics Theses | Department of Mathematics and Statistics

12-11-2012

# A Comparative Study Between Genotypes and Ages of Eyes Using Morphometric Measures of Retinal Pigment Epithelial Cells

Micheal Shola Folarinde
mfolarinde@gmail.com

Follow this and additional works at: http://digitalarchive.gsu.edu/math_theses

Recommended Citation
Folarinde, Micheal Shola, "A Comparative Study Between Genotypes and Ages of Eyes Using Morphometric Measures of Retinal Pigment Epithelial Cells" (2012). *Mathematics Theses.* Paper 117.

This Thesis is brought to you for free and open access by the Department of Mathematics and Statistics at Digital Archive @ GSU. It has been accepted for inclusion in Mathematics Theses by an authorized administrator of Digital Archive @ GSU. For more information, please contact digitalarchive@gsu.edu.

# A COMPARATIVE STUDY BETWEEN GENOTYPES AND AGES OF EYES USING MORPHOMETRIC MEASURES OF RETINAL PIGMENT EPITHELIAL CELLS

by

MICHEAL S FOLARINDE

UNDER THE DIRECTION OF DR. YI JIANG

ABSTRACT

Aged-related macular degeneration (AMD) is a common eye condition among people older than 65 years and is a leading cause of vision loss. It gradually destroys the macula, the part of the eye that provides sharp, central vision needed for seeing objects clearly. This study aims to test the hypothesis that the morphology of retina pigment epithelium, a key site of AMD pathology, can reflect the various stresses aging and AMD progression impose. We first identify and separate the young and old age group for mouse eyes. Then we classify, the mouse eyes using two genotypes (C57BL/6L, RD10), and two age group (young, old).We show that without dimensional reduction, the cell area and shape measures do not provide good classification of the mouse eyes.

But with the dimension reduction at the eye level, the cell area and shape measures provide excellent classification for mouse genotype and age.

INDEX WORDS: Retina pigment epithelium, C57BL/6L, RD10, Morphometric

# A COMPARATIVE STUDY BETWEEN GENOTYPES AND AGES OF EYES USING MORPHOMETRIC MEASURES OF RETINAL PIGMENT EPITHELIAL CELLS

by

MICHEAL S FOLARINDE

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2012

Copyright by
Micheal S. Folarinde
2012

A COMPARATIVE STUDY BETWEEN GENOTYPES AND AGES OF EYES USING MORPHOMETRIC MEASURES OF RETINAL PIGMENT EPITHELIAL CELLS

by

MICHEAL S FOLARINDE

Committee Chair: Dr. Yi Jiang

Committee Dr. Xin Qi

Dr. Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2012

## DEDICATION

I would like to dedicate this thesis to the glory of Almighty God who gave me the strength and the ability to go through the rigor of school. This work is also dedicated to my wife, Bunmi Folarinde and my daughters, Blessing Folarinde and Busayo Folarinde, who encouraged and supported me throughout the course of this study.

## ACKNOWLEDGEMENTS

I thank my advisor, Dr. Yi Jiang, who with her busy schedule took me as her student. Under her supervision and support, I was able to learn more about Statistics, research techniques, and developed programming skills that would help me with my future career. I am proud to be one of Dr. Yi Jiang students. Also, I would like to appreciate Dr. Yuanhui Xiao and Dr. Xin Qi who accepted to guide and advise through my data analysis. I would like to appreciate my motivator in the department, Dr. Yichuan Zhao who gave me some insights on how to work on thesis research.

Lastly, I appreciate my wife, Bunmi Folarinde, my daughters, Blessing and Busayo Folarinde, friends and colleagues who supported me in various ways thoughtout my study period.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....v

LIST OF TABLES ..... .....ix

LIST OF FIGURES.....x

1 INTRODUCTION .....1

1.1 Purpose of the Study .....1

1.2 Study Data.....3

2 DATA EXPLORATION .....4

2.1 Data Exploration .....4

2.2 Data Exploration for Genotype C57BL/6J .....5

2.3 Data Exploration for Genotype RD10.....8

3 RESEARCH METHODOLOGY CLASSIFICATION with MORPHOMETRIC MEASURE OF CELLS. 11

3.1 Linear Discriminant Analysis of Genotype with Morphometric of Cells .....11

3.2 Quadratic Discriminant Analysis of Genotype with Morphometric Cells.....12

3.3 Linear Discriminant Analysis of Age Group with Morphometric Cells.....13

3.4 Quadratic Discriminant Analysis of Age Group with Morphometric Cells.....14

4 RESEARCH METHODOLOGY CLASSIFICATION WITH ONE MORPHOMETRIC MEASURE OF EACH EYES.....15

4.1 Linear Discriminant Analysis of Genotype for each eye with area.....15

4.2 Quadratic Discriminant Analysis of Genotype for each eye with area.....16

4.3 Linear Discriminant Analysis of Age group for each eye with area.....17

**4.4 Quadratic Discriminant Analysis of Age group for each Eyes with area……..…18**

**4.5 Linear Discriminant Analysis of Genotype for each eye with eccentricity……..19**

**4.6 Quadratic Discriminant Analysis of Genotype for each eye with eccentricity……20**

**4.7 Linear Discriminant Analysis of Age group for each eye with eccentricity ……..21**

**4.8 Quadratic Discriminant Analysis of Age group for each eyes with eccentricity……..……22**

**5 RESEARCH METHODOLOGY CLASSIFICATION WITH TWO MORPHOMETRIC MEASURE OF EACH EYES……..…273**

**5.1 Linear Discriminant Analysis of Genotype with Statistics Morphometric Measure of Cell for each eyes……..……23**

**5.2 Quadratic Discriminant Analysis of Genotype with Statistics Morphometric of Cells for each eyes……24**

**5.3 Linear Discriminant Analysis of Age group with Statistics Morphometric Measure of Cells Of Each Eyes……..25**

**5.4 Quadratic Discriminant Analysis of Age group with Statistics Morphometric Measure of Cells of Each Eyes……26**

**6 Conclusion……..28**

**REFERENCES……30**

**APPENDICES .....31**

**Appendix A Sas code for Data Cleaning……31**

**Appendix B Sas code for Density Curve of Genotype C57BL/6J……..45**

**Appendix C Sas Code For Density Curve of Genotype RD10……...47**

**Appendix D Sas Code For Linear /Quadratic Discriminant function of Morphometric Measure Cells ……..…49**

**Appendix E Sas Code For Linear /Quadratic Discriminant function of Statistics Morphometric Measure Cells……..……..50**

**Appendix F Sas Output for Linear Discriminant function for Genotype ……52**

**Appendix G Sas Output for Qaudratic Discriminant function for Genotype …….……..…56**

**Appendix H Sas Output for Linear Discriminant function for Age Group ……..59**

**Appendix I Sas Output for Qaudratic Discriminant function for Age Group ……..62**

**Appendix J Sas Output for Linear Discriminant function for Genotype with Statistics Morphometric Measure Of each Eyes …….……..……..65**

**Appendix K Sas Output for Qaudratic Discriminant function for Genotype with Statistics Measure Of each Eyes ……69**

**Appendix L Sas Output for Linear Discriminant function for Age Group with Statistics Measure Of each Eyes ……72**

**Appendix M Sas Output for Linear Discriminant function for Age Group with Statistics Morphometric Measure Of each Eyes ……67**

**Appendix N SAS Output for Validated and training data of cells……78**

## LIST OF TABLES

**Table 1.1 Summary Statistics Genotype of RD10…….3**

**Table 1.2 Summary Statistics Genotype of CB56L/6L…….4**

**Table 2.1 Summary of Cumulative proportion Area and Eccentricity cell of CB56L/6L…….7**

**Table 2.2 Summary of Cumulative proportion Area and Eccentricity cell of RD10……10**

**Table 3.1 Number of Observations and Percent Classified into genotype……12**

**Table 3.2 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis……12**

**Table 3.2.1 Group Classification for Genotype group……...13**

**Table 3.3 Number of Observations and Percent Classified into Age group Linear Discriminant Function ……13**

**Table 3.4 Number of Observations and Percent Classified into Age group using Quadratic Linear Determinant……...14**

**Table 3.4.1 Group Classification for Age group……14**

**Table 4.1 Number of Observations and Percent Classified into genotype using Linear Determinant Analysis (Area)……15**

**Table 4.2 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis (Area)……16**

**Table 4.3 Number of Observations and Percent Classified into Age Group using Linear Determinant Analysis (Area)……17**

**Table 4.4 Number of Observations and Percent Classified into Age group using Quadratic Determinant Analysis (Area)……..18**

**Table 4.5 Number of Observations and Percent Classified into genotype using Linear Determinant Analysis (Eccentricity)……19**

**Table 4.6 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis (Eccentricity)……..20**

**Table 4.7 Number of Observations and Percent Classified into Age Group using Linear Determinant Analysis (Eccentricity)……21**

**Table 4.8 Number of Observations and Percent Classified into Age group using Quadratic Determinant Analysis (Eccentricity)……22**

**Table 5.1 Number of Observations and Percent Classified into genotype using Linear Determinant Analysis (five summary statistics) ……23**

**Table 5.2 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis (five summary statistics)……...24**

**Table 5.3 Number of Observations and Percent Classified into Age Group using Linear Determinant Analysis (five summary statistics)……..…25**

**Table 5.4 Number of Observations and Percent Classified into Age group using Quadratic Determinant Analysis (five summary statistics)…….....26**

**Table 5.5 Group Classification for Genotype group (Five Summary Statistics)…….....27**

**Table 5.6 Group Classification for Age group (Five Summary Statistics)…….....28**

## LIST OF FIGURES

**Figure 1.1 Picture of the Eyes……3**

**Figure 2.1 Density Curve Area Shape cell Age 320 of CB57BL/6J…….5**

**Figure 2.2 Density Curve Area Shape cell Age 180 of CB57BL/6J……..6**

**Figure 2.3 Density Curve Area Shape cell Age 70 of CB57BL/6J……6**

**Figure 2.4 Density Curve Area Shape cell Age 32 of RD10……8**

**Figure 2.5 Density Curve Area Shape cell Age 70 of RD10……8**

**Figure 2.6 Density Curve Area Shape cell Age 180 of RD10……9**

# 1 INTRODUCTION

## 1.1 Purpose of the Study

Age-related macular degeneration (AMD) is a leading cause of vision loss in adult 60 years of age and older. AMD affects the macula, the part of eye that allows you to see fine detail, and gradually destroy sharp, central vision. Central vision is needed for seeing objects clearly and for common daily tasks such as reading and driving. In our aging society, AMD is a looming epidemic. Presently there is no effective treatment for this disease. AMD develops differently in different people. In some cases, AMD advances so slowly that patient notice little change in their vision. In others, the disease progresses faster and may lead to a loss of vision in both eyes.

Different types of AMD development require different treatments. But we cannot yet predict which patient will develop which type of AMD. There is a great need for diagnostic and prognostic tools for AMD.

In the eye, retina pigment epithelium (RPE) is a key site of AMD pathology. Different processes associated AMD progression would impose different stresses on RPE cells. Based on previous work of Jiang et al(Jiang et al. 2012), we hypothesize that the morphology of retinal pigment epithelium can be related to AMD disease progression, and potentially be used as a diagnostic, even prognostic, indicator for AMD.

This thesis project is aimed toward testing this hypothesis. We use mouse RPE morphology data to discriminate the corresponding eye's age and genotype. The focus is on the comparative study between age and genotype classification with and without dimensional reduction at the eye level.

We have access to large number of mouse eye data, which include genotype and age of the eyes, and morphometric measures of each cell identified within the eyes of a mouse. We focus on

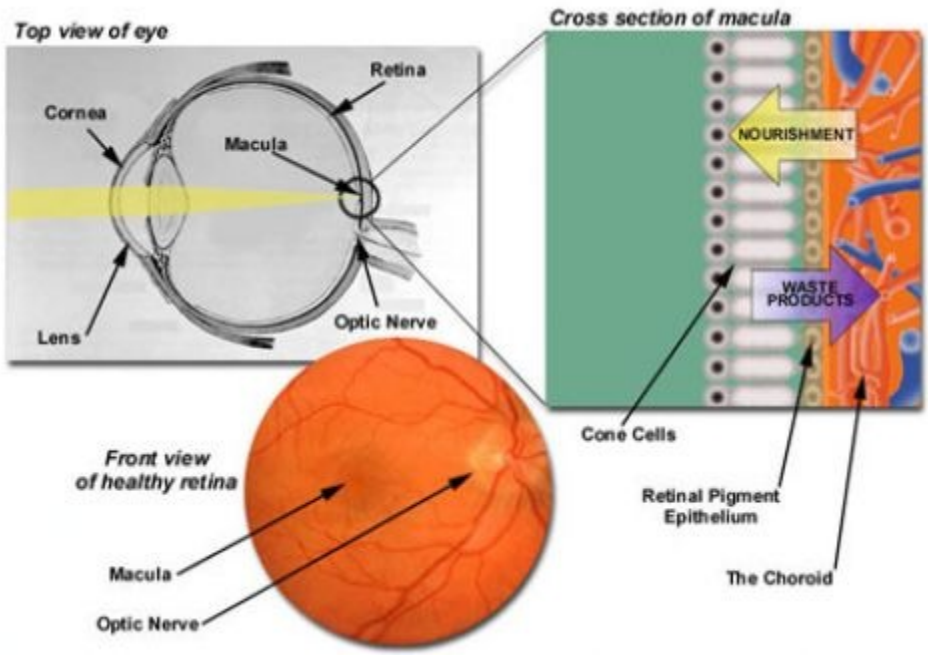


**Figure 1 - Picture of the eyes showing Dry and Wet Macular Degeneration**

Source: http://www.your-eye-sight.org/cause-of-macular-degeneration.html

- Two genotypes (C57BL/6J, RPE65) sample of 110 Eyes
- Seven age groups (P30, P45, P60-70, P90-110, P180, P360, P720)
- Two morphometric measures (eccentricity, area) sample of 199803 cells.

One of the purposes of study is to identify the best point of separation of young and old for mouse. After identifying the point of separation of young and old, we will compare two ways of classification, based on cell level data for genotype and age using cell size and shape. The goal is to find the relationship between the RPE patterns to AMD progression, and help to understand and predict AMD progression. To achieve the above linear discriminant analysis (LDA) and quadratic discriminant analysis(QDA) are used to analyze the data.

## 1.2 Study Data

The study data contains one hundred and ten (110) individual eyes with two types of genotype Rd10 and C57BL/6J. C57BL/6J is considered the wild type and RD10 is a mutant with a deletion in the RPE related gene.

These eyes have different ages which range from post natal 300 days (p30) to 720 days (2 years). Image analysis using cell profiler offer 21 morphometric measures for each the number of cells measured varies according to individual eyes. Total of 199,803 individual cells were measured for this study.

The data on genotype and age as it related to area and eccentricity will be properly classified and segregated. Table 1.1 contain mean and cumulative mean of Rd10 with respect to area and eccentricity while Table 1.2 contain mean and cumulative mean of CB57BL/6J.

The age was properly segregated into young and old group using density curve and cumulative proportion of data pattern. The variance between the genotype will be uncovered. The difference between methods utilized for the study will clarify.

**Table 1.1 Summary Statistics of Genotype RD10**

| AGE | AreaShape-area | Cummulative |
|---|---|---|
| 30 | 144.55 | 144.55 |
| 45 | 159.62 | 304.17 |
| 60 | 171.67 | 475.84 |
| 61 | 156.38 | 632.22 |
| 100 | 161.28 | 793.5 |
| 180 | 160.04 | 953.54 |
| 330 | 198.43 | 1151.97 |
| 723 | 182.56 | 1334.53 |
| 732 | 175.95 | 1510.48 |

**Table 1.2 Summary Statistics of Genotype CB57/BL6L**

| AGE | AreaShape-area | Cummulative |
|---|---|---|
| 30 | 144.55 | 144.55 |
| 45 | 159.62 | 304.17 |
| 60 | 171.67 | 475.84 |
| 61 | 156.38 | 632.22 |
| 100 | 161.28 | 793.5 |
| 180 | 160.04 | 953.54 |
| 330 | 198.43 | 1151.97 |
| 723 | 182.56 | 1334.53 |
| 732 | 175.95 | 1510.48 |

# CHAPTER 2

## 2 DATA EXPLORATION AND SAMPLING

### 2.1 Data Exploration

We examine the pattern of specific individual age with similar genotype with respect to area and shape of cells within the eyes. We plot the density Curve using the mean of individual eyes. The summary of the cumulative frequency proportion is examined.

### 2.2 Data Exploration for Genotype C57BL/6J

SAS code is used to generate the density curve using the average of area shape cell with respect to Individual eye and genotype. The Figure 2.1 below is the density curve of area of age 320 with Genotype C57BL/6J. The Black curve represent age less than 320 while the Blue curve is age greater than 320. Next is Figure 2.2 is the density curve of area of age 400 Genotype C57BL/6J.The Black curve represent age less than 400 while the Blue curve is age greater than 400. Note that the Y-axis represent the density while the X-axis is the area (Value).

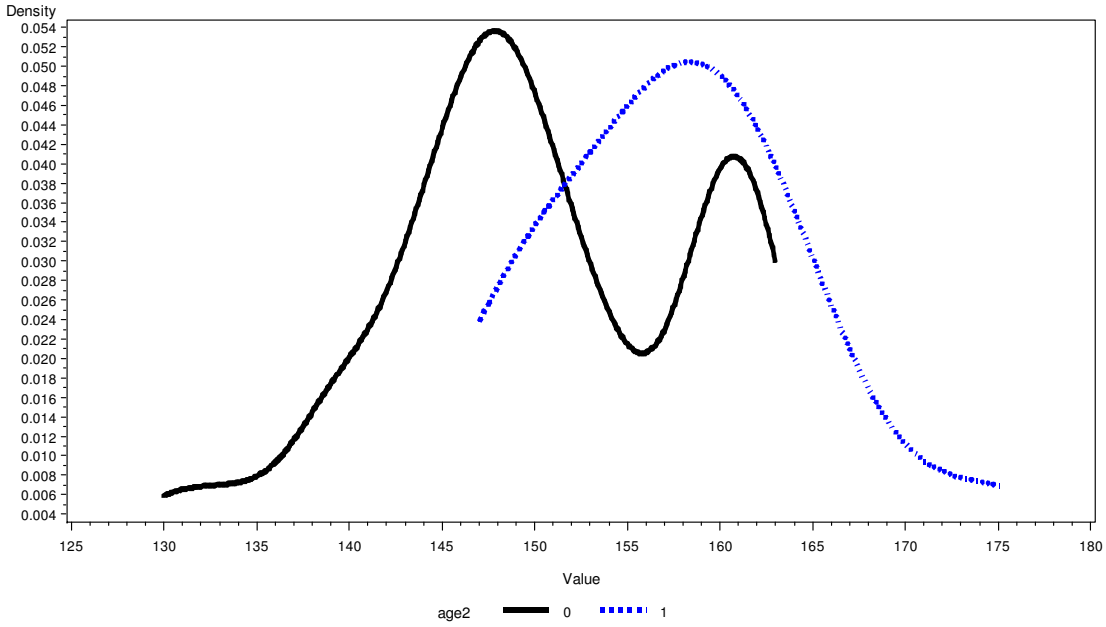


**Figure 2.1 Density Curve Area shape cell of Age 320 of C57BL/6J**

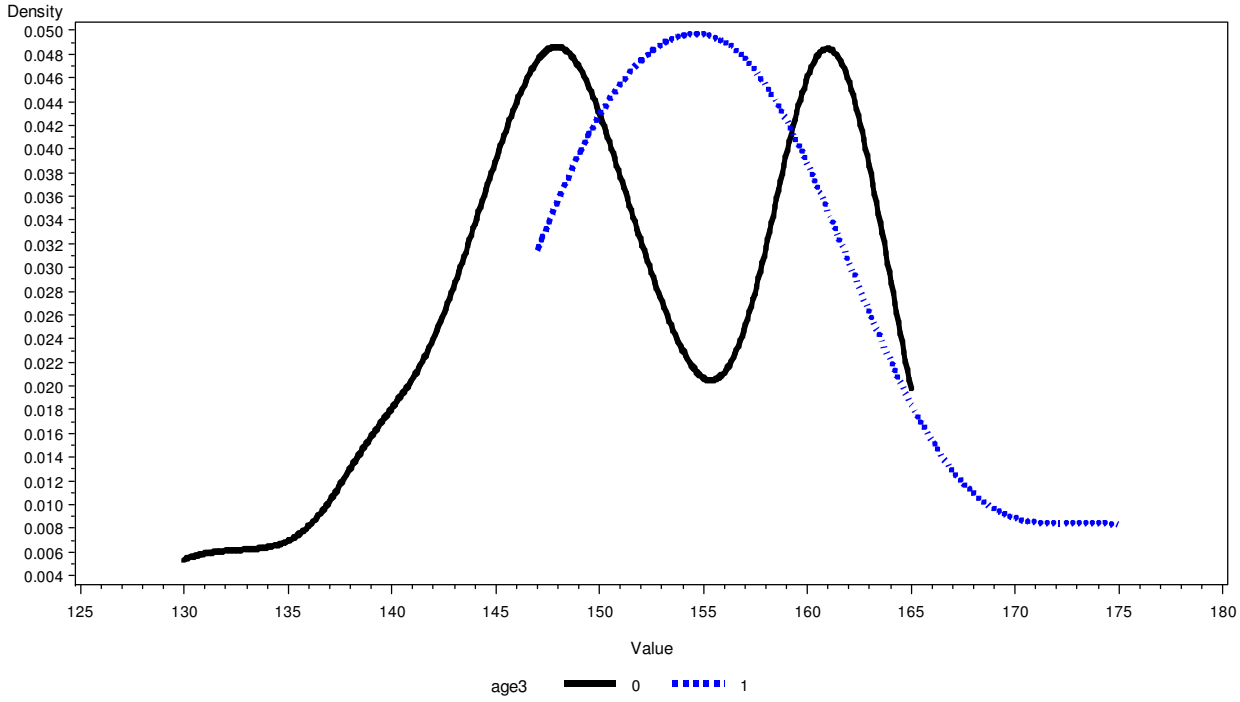


**Figure 2.2 Density Curve Area cell of Age 400 of C57BL/6J**

Below is Figure 2.3 is the density curve of area of age 70 Genotype C57BL/6J .The Black curve represent age less than 70 while the Blue curve is age greater than 70.

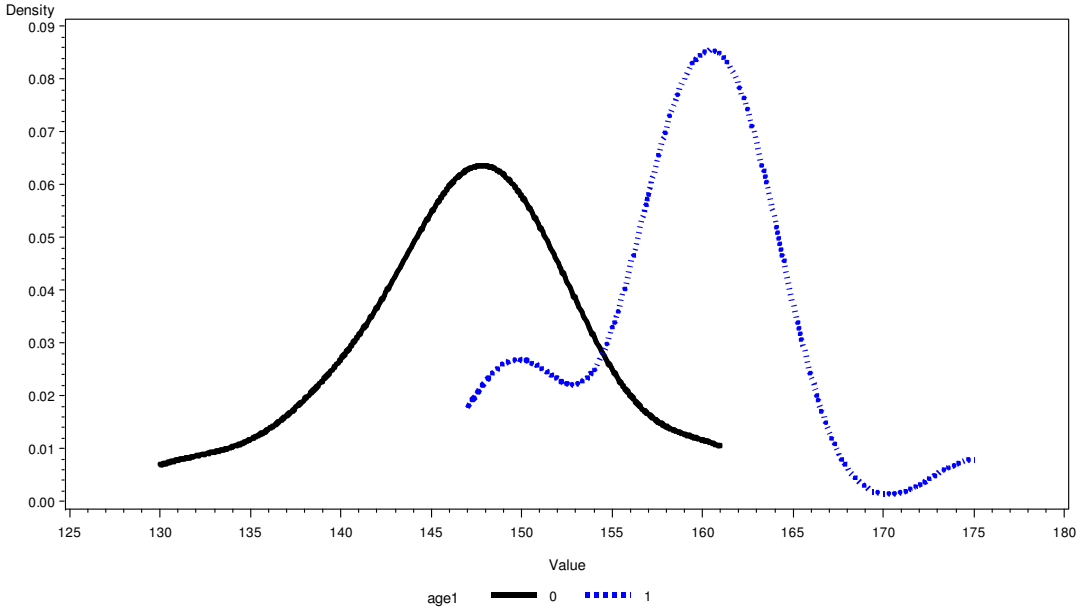


**Figure 2.3 Density Curve Area shape cell of Age 70 of C57BL/6J**

The pattern of density Curve of Figure 2.1 (age 320) and Figure 2.2 (age 400) for age less than 320 and age less than 400 look the similar with a lots of overlap. The density curve for age less than 70 and age greater than 70 will make a good separation because there are minimum overlaps. Below is Table 2.1 depicts the cumulative proportion Area and Eccentricity cell of each eye for Genotype CB57BL/6J, which shows my age 180 as my cut off point of 50 percent.

**Table 2.1 Summary of Cumulative Proportion Area and Eccentricity cell**

| AGE | CB57BL/6J | |
|---|---|---|
| | Area<br><br>Cummulative Proportion | Eccentricity<br><br>Cummalative Proprotion |
| 30 | 0.104771219 | 0.109965636 |
| 45 | 0.209660586 | 0.218213058 |
| 60 | 0.316394379 | 0.326460481 |
| 61 | 0.424145559 | 0.436426117 |
| 180 | 0.538939685 | 0.546391753 |
| 330 | 0.653523771 | 0.661512027 |
| 700 | 0.76404487 | 0.773195876 |
| 720 | 0.885744104 | 0.888316151 |
| 722 | 1 | 1 |

## 2.3 Data Exploration for Genotype RD10

The Figure 2.4 below is the density curve of area of age 320 with Genotype RD10. The black curve represent age less than 320 while the blue curve is age greater than 320.Note for all the density Curve the Y-axis represent the density while X-axis represent the area (Value). The pattern of density Curve of RD10 Figure 2.5 (age 70) and Figure 2.6 (age 400) look the same but are different. The density curve look like a bell shape for age less than 70 but age greater than 70 and this will make a good separation with a minimum overlap over other density curve. Based on the graph pattern of the two genotype, age 70 will be point of separation for young and old.

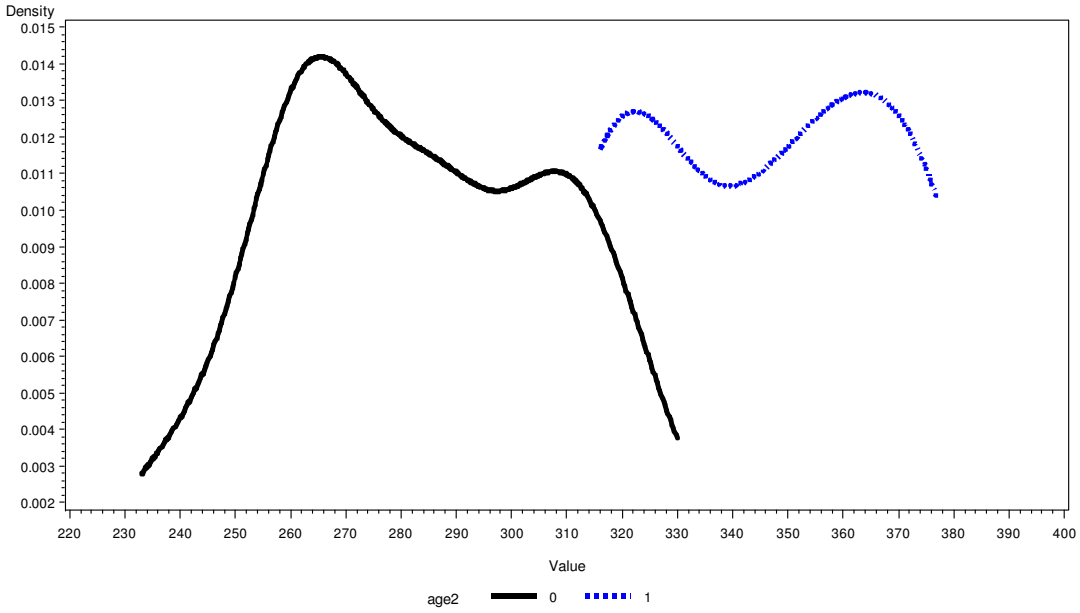


**Figure 2.4 Density Curve Area cells of Age 320 of RD10**

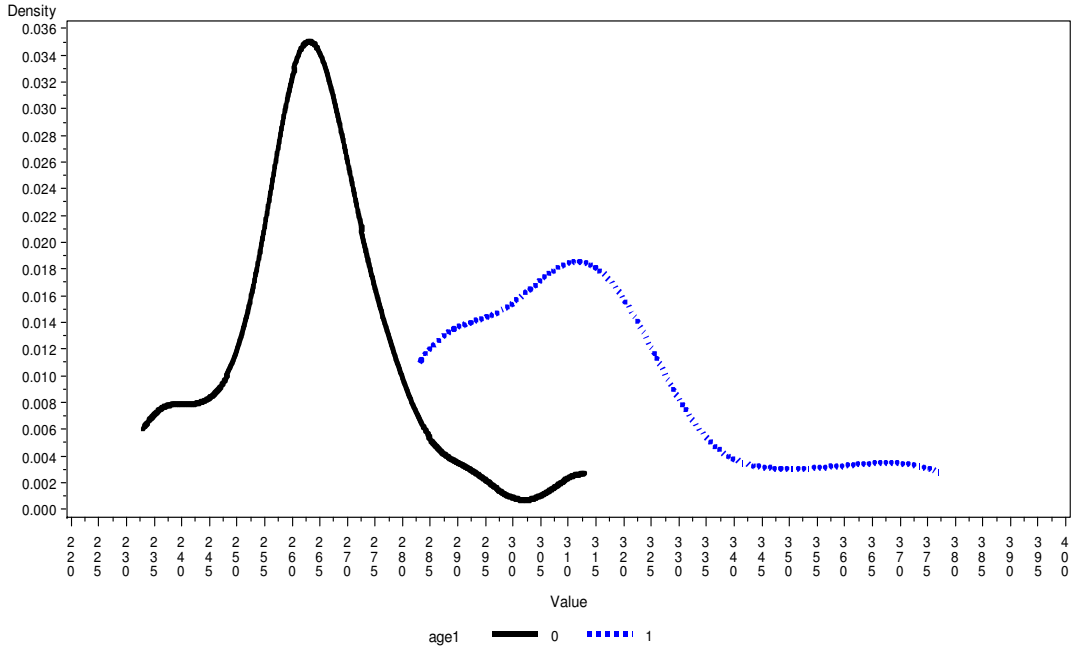


**Figure 2.5 Density Curve Area shape cell of Age 70 of RD10**

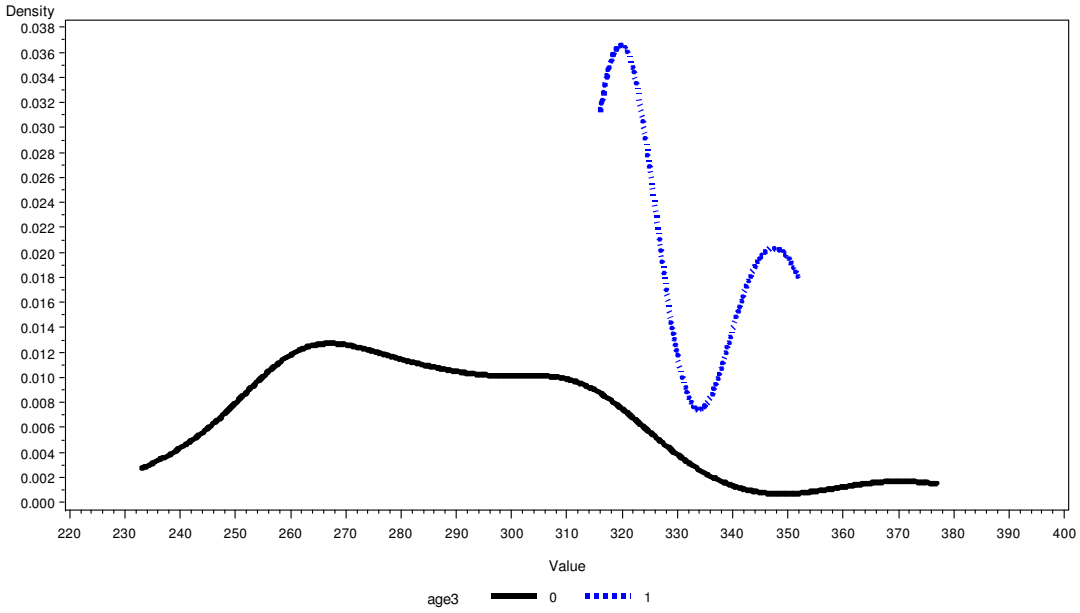


**Figure 2.6 Density Curve Area shape cell of Age 400**

The next table 2.1 depicts the cumulative proportion Area and Eccentricity cell of each eye for Genotype RD10, which shows my age 100 as my cut off point of 50 percent.

**Table 2.2 Summary of Cumulative Proportion Area and Eccentricity cell**

| Age | **RD10** | |
|---|---|---|
| | **Area Cummulative Proportion** | **Eccentricity Cummulative Proportion** |
| 30 | 0.095698056 | 0.102167183 |
| 45 | 0.201373073 | 0.213622291 |
| 60 | 0.315025687 | 0.321981424 |
| 61 | 0.418555691 | 0.42879257 |
| 100 | 0.525329697 | 0.540247678 |
| 180 | 0.631282771 | 0.656346749 |
| 330 | 0.762651607 | 0.775541796 |
| 723 | 0.88351385 | 0.886996904 |
| 732 | 1 | 1 |

Based on Density curve and the cumulative proportion table above my cut of age is age 70.Eyes with age less than 70 is classified as young while age of eyes greater than 70 classified as old.

# 3 METHOD FOR CLASSIFICATION WITH MORPHOMETRIC MEASURE OF CELLS

One of the purposes of this thesis research is to correctly predict the classification of morphometric of cell of eyes into age and genotype using linear discriminant analysis and quadratic linear analysis. The Morphometric cell measures under consideration are area and eccentricity measure of eyes. Procedure begins with a set of observations where both group membership and the values of the interval variables are known. The end result of the procedure is a model that allows prediction of group membership when only the interval variables are known.

## 3.1 Linear Discriminant Analysis of Genotype with Morphometric Measure of Cells.

The total of all the genotype with the total morphometric of each cell is analysis. We have total of 199804 individual cells. The dependent variable which is the Genotype group is classified into CB57L/6J (0) and RD10 (1) while the predictor or independent variable is area and eccentricity. The linear discriminant analysis (LDA) of the above data set is done assumed equal variance. The pooled covariance matrix is 4.52443 while the overall error rate is .4113.

**Table 3.1 Number of Observations and Percent Classified into genotype**

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 64348<br>54.69 | 53311<br>45.31 | 117659<br>100 |
| 1 | 30344<br>36.94 | 51801<br>63.06 | 82145<br>100 |
| Total | 94692<br>47.39 | 105112<br>52.61 | 199804<br>100 |
| Priors | 0.5 | 0.5 | |

From Table 3.1 above we see 199804 cells observation of which 54.69 percent were correctly classified as Genotype CB57L/6J (0) better than 52.61 percent Genotype RD10 Classified correctly

### 3.2 Quadratic Discriminant Analysis of Genotype with Morphometric Measure of Cells

Using, the total size of 199804 individual cells, the dependent variable which is the group is classified into CB57L/6J and RD10 while the predictor or independent variable is area and eccentricity. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance matrix for CB57L/6J and RD10 are 4.47985 and 4.58455 respectively.

**Table 3.2 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis**

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| **0** | 65808<br>55.93 | 51851<br>44.07 | 117659<br>100 |
| **1** | 30757<br>37.44 | 51388<br>62.56 | 82145<br>100 |
| **Total** | 96565<br>48.33 | 103239<br>51.67 | 199804<br>100 |
| **Priors** | 0.5 | 0.5 | |

From Table 3.2 above we see 199804 cells observation of which 62.56 percent genotype RD10 (1) Classified correctly while 55.93 percent were correctly classified as genotype CB57L/6J (0). Overall, 40.76% of the observations were misclassified.

**Table 3.2.1 Group Classification for Genotype Group for Cells**

| Class Level Information | | | | | |
|---|---|---|---|---|---|
| **genotype1** | **Variable Name** | **Frequency** | **Weight** | **Proportion** | **Prior Probability** |
| **0** | **_0** | **117659** | **117659** | **0.588872** | **0.500000** |
| **1** | **_1** | **82145** | **82145** | **0.411128** | **0.500000** |

Table 3.2.1 shows that the CB57L/6J (0) contributes most to Genotype group separations which has 57.8 percent of total cell 199804.

### 3.3 Linear Discriminant Analysis of Age group with Morphometric Measure of Cells.

The total of all the genotype with the total Morphometric of each cell is analysis using linear discriminant Analysis (LDA). We have total size of 199804 individual cells. The dependent variable which is the Age group is classified into young (0) and old (1) while the predictor or independent variable is area and eccentricity. The linear discriminant analysis(LDA) of the above data set is done assumed equal variance. The pooled covariance matrix is 4.55498 while the error rate is .4540.

**Table 3.4 Number of Observations and Percent Classified into Age group Linear Discriminant Function**

| Age Group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 69023<br>59.72 | 46557<br>40.28 | 115580<br>100 |
| 1 | 42552<br>50.52 | 41672<br>49.48 | 84224<br>100 |
| Total | 111575<br>55.84 | 88229<br>44.16 | 199804<br>100 |
| Priors | 0.5 | 0.5 | |

From Table 3.3 above we see 199804 cells observation of which 59.72 percent young group (0) classified correctly which is better than 49.48 percent correctly classified as older group(1). Overall, 45.40% of the observations were mis-classified.

### 3.4 Quadratic Discriminant Analysis of Age group with respect to Morphometric Measure of Cells.

Using, the total size of 199804 individual cells, the dependent variable which is the Age group is classified into young (0) and old (1) while the predictor or independent variable is area and eccentricity. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance Matrix for CB57L/6J and RD10 are 4.31550 and 4.82505 respectively error rate 0.4477.

**Table 3.4 Number of Observations and Percent Classified into Age group using Quadratic Linear Determinant**

| Age group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 93039<br>80.5 | 22541<br>19.5 | 115580<br>100 |
| 1 | 58913<br>69.95 | 25311<br>30.05 | 84224<br>100 |
| Total | 151952<br>76.05 | 47852<br>23.95 | 199804<br>100 |
| Priors | 0.5 | 0.5 | |

From Table 3.4 above we see 199804 cells observation of which 80.5 percent young group (0) classified correctly which is better than 30.05 percent correctly classified as older group(1). Overall, 44.73% of the observations were misclassified.

**Table 3.4.1 Group Classification for Age Group**

| Class Level Information | | | | | |
|---|---|---|---|---|---|
| age1 | Variable Name | Frequency | Weight | Proportion | Prior Probability |
| 0 | _0 | 115580 | 115580 | 0.578467 | 0.500000 |
| 1 | _1 | 84224 | 84224 | 0.421533 | 0.500000 |

Table 3.4.1 shows that the young group (0) contributes most to age group separations which has 57.8 percent of total cell 199804

Note: classifying the cells data into training and Validation data set then classifying genotype and age using LDA and QDA produces a higher misclassification rate and invariably higher overall error rate. See Appendix-N for output of each classification.

# 4 RESEARCH METHODOLOGY CLASSIFICATION WITH ONE MORPHOMETRIC MEASURE OF EACH EYES

The second of the purposes of this thesis research is to correctly predict the classification of Morphometric measure of cells of each eye into age and genotype using linear discriminant analysis and quadratic linear analysis. The Morphometric cell measures under consideration are 5 percentile Area , 25 percentile Area , 50$^{th}$ percentile Area, 75 percentile Area , 95 percentile Area , 5 percentile Eccentricity 25 percentile Eccentricity 50 percentile Eccentricity 75 percentile Eccentricity and 95 percentile Eccentricity measure of each eyes. Area and Eccentricity percentile are considered separately.

## 4.1 Linear Discriminant Analysis of Genotype for each Eye with Area

The five summary statistics of 110 individual eyes are an*a*lyzed using linear discriminant analysis. The dependent variable which is the genotype group is classified into CB57L/6J (0) and RD10 (1) while the predictor or independent variable is five numbers Summary of area only. The pooled covariance Matrix is 14.923 while the error rate is .0633. From Table 4.1 above we see 110 individual eyes 94 percent Genotype CB57L/6J (0) were correctly classified as better than 93.33 percent Genotype RD10 Classified correctly. Data could not be divided into train and validation data because of size of eye data.

**Table 4.1 Number of Observations and Percent Classified into genotype using Linear Determinant Analysis (Area)**

| Genotype | 0 | 1 | Total |
|---|---|---|---|
| **0** | 47<br>94 | 3<br>6 | 50<br>100 |
| **1** | 4<br>6.67 | 56<br>93.33 | 60<br>100 |
| **Total** | 51<br>46.36 | 59<br>53.64 | 110<br>100 |
| **Priors** | 0.5 | 0.5 | |

### 4.2 Quadratic Discriminant Analysis of Genotype for each Eye with Area

The five summary statistics of 110 individual eyes are analyzed. The dependent variable which is the group is classified into CB57L/6J and RD10 while the predictor or independent variables are five number summary of area only. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance Matrix for CB57L/6J and RD10 are 14.849 and 13.083 respectively while the error rate 0.0633.

**Table 4.2 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis (AREA)**

| Genotype | 0 | 1 | Total |
|---|---|---|---|
| **0** | 47<br>94 | 3<br>6 | 50<br>100 |
| **1** | 4<br>6.67 | 56<br>93.33 | 60<br>100 |
| **Total** | 51<br>46.36 | 59<br>53.65 | 110<br>100 |
| **Priors** | 0.5 | 0.5 | |

From table 4.2, we discovered that the results are the same.

### 4.3 Linear Discriminant Analysis of Age Group for each Eye with Area

The five summary statistics of 110 individual eye which will be analyzed using linear discriminant analysis. The dependent variable which is the group is classified into young (0) and old (1) while the predictor or independent variable are five number summary of area shape . The Linear Discriminant analysis of the above data set is done assumed equal variance. The pooled covariance Matrix is 15.296 while the error rate is 0.0817.

From Table 4.3 above we see 110 individual eyes of which 92.59 percent young group (0) classified correctly which is better than 91.07 percent correctly classified as older group(1). Overall, 8.17% of the observations were mis-classified.

**Table 4.3 Number of Observations and Percent Classified into Age Group using Linear Determinant Analysis (Area)**

| Age Group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>92.59 | 4<br>7.41 | 54<br>100 |
| 1 | 5<br>8.93 | 51<br>91.07 | 56<br>100 |
| Total | 55<br>50 | 55<br>50 | 110<br>100 |
| Priors | 0.5 | 0.5 | |

### 4.4 Quadratic Discriminant Analysis of Age Group for each Eye with Area

The five summary statistics of 110 individual eyes are analyzed. The dependent variable which is the group is classified into young (0) and old (1) while the predictor or independent variable are five number summary of area only. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance matrix for are young (0) and old (1) 12.5155 and 15.51651 respectively while the error rate 0.0731.

From Table 4.4 above we see 110 individual eyes of which 90.74 percent young group (0) classified correctly which is not better than 94.64 percent correctly classified as older group(1).

**Table 4.4 Number of Observations and Percent Classified into Age group using Quadratic Determinant Analysis (Area)**

| Age group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 49<br>90.74 | 5<br>9.26 | 54<br>100 |
| 1 | 3<br>5.36 | 53<br>94.64 | 56<br>100 |
| Total | 52<br>47.27 | 58<br>52.73 | 110<br>100 |
| Priors | 0.5 | 0.5 | |

### 4.5 Linear Discriminant Analysis of Genotype for each Eye with Eccentricity

The five summary statistics of 110 individual eyes are an*a*lyzed using linear discriminant analysis. The dependent variable which is the genotype group is classified into CB57L/6J (0) and RD10 (1) while the predictor or independent variable is five numbers Summary of Eccentricity only. The pooled covariance Matrix is -45.26 while the error rate is .15. From Table 4.1 above we see 110 individual eyes 100 percent Genotype CB57L/6J (0) were correctly classified as better than 70percent Genotype RD10 Classified correctly.

**Table 4.5 Number of Observations and Percent Classified into genotype using Linear Determinant Analysis (Eccentricity)**

| Genotype | 0 | 1 | Total |
|---|---|---|---|
| **0** | 50<br>100 | 0<br>0.00 | 50<br>100 |
| **1** | 18<br>30 | 42<br>70 | 60<br>100 |
| **Total** | 68<br>61.82 | 42<br>38.18 | 110<br>100 |
| **Priors** | 0.5 | 0.5 | |

### 4.6 Quadratic Discriminant Analysis of Genotype for each Eye with Eccentricity

The five summary statistics of 110 individual eyes are analyzed. The dependent variable which is the group is classified into CB57L/6J and RD10 while the predictor or independent variables are five number summary of Eccentricity only. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance Matrix for CB57L/6J and RD10 are -47.85 and -50.24 respectively while the error rate 0.07.

From Table 4.6 below we see 110 individual eyes 86 percent Genotype CB57L/6J (0) were correctly classified as less better than 100 percent Genotype RD10 Classified correctly.

**Table 4.6 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis (Eccentricity)**

| Genotype | 0 | 1 | Total |
|---|---|---|---|
| **0** | 43<br>86 | 7<br>14 | 50<br>100 |
| **1** | 0<br>0.00 | 60<br>100 | 60<br>100 |
| **Total** | 43<br>39.09 | 67<br>60.91 | 110<br>100 |
| **Priors** | 0.5 | 0.5 | |

### 4.7 Linear Discriminant Analysis of Age Group for each Eye with Eccentricity

The five summary statistics of 110 individual eye which will be analyzed using linear discriminant analysis. The dependent variable which is the group is classified into young (0) and old (1) while the predictor or independent variables are five number summary of shape. The Linear Discriminant analysis of the above data set is done assumed equal variance. The pooled covariance Matrix is 15.296 while the error rate is 0.2824.

From Table 4.7 below we see 110 individual eyes of which 68.52 percent young group (0) classified correctly which is less better than 75 percent correctly classified as older group(1). Overall, 28.24 of the observations were mis-classified.

**Table 4.7 Number of Observations and Percent Classified into Age Group using Linear Determinant Analysis (Eccentricity)**

| Age Group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 37<br>68.52 | 17<br>31.48 | 54<br>100 |
| 1 | 14<br>25 | 42<br>75 | 56<br>100 |
| Total | 51<br>46.36 | 59<br>53.64 | 110<br>100 |
| Priors | 0.5 | 0.5 | |

### 4.8 Quadratic Discriminant Analysis of Age Group for each Eye with Eccentricity

The five summary statistics of 110 individual eyes are analyzed. The dependent variable which is the group is classified into young (0) and old (1) while the predictor or independent variable are five number summary of eccentricity only. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance matrix for are young (0) and old (1) -53.09609 and -43.352 respectively while the error rate 0.2156.

From Table 4.8 above we see 110 individual eyes of which 92.59 percent young group (0) classified correctly which is better than 64.29 percent correctly classified as older group(1).

**Table 4.8 Number of Observations and Percent Classified into Age group using Quadratic Determinant Analysis (Eccentricity)**

| Age group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>92.74 | 4<br>7.41 | 54<br>100 |
| 1 | 20<br>35.71 | 36<br>64.29 | 56<br>100 |
| Total | 70<br>47.27 | 40<br>52.73 | 110<br>100 |
| Priors | 0.5 | 0.5 | |

# 5 RESEARCH METHODOLOGY CLASSIFICATION WITH STATISTICS MORPHOMETRIC MEASURE OF EACH EYES

The second of the purposes of this thesis research is to correctly predict the classification of Morphometric measure of cells of each eye into age and genotype using linear discriminant analysis and quadratic linear analysis. The Morphometric cell measures under consideration are 5 percentile Area , 50 percentile Area , 25$^{th}$ percentile Area, 75 percentile Area , 95 percentile Area , 5 percentile Eccentricity 25 percentile Eccentricity 50 percentile Eccentricity 75 percentile Eccentricity and 95 percentile Eccentricity measure of each eyes. Percentiles of area and eccentricity are considered together.

## 5.1 Linear Discriminant Analysis of Genotype with Statistics Morphometric measure of Cell for each eyes.

The five summary statistics of 110 individual eyes are an*a*lyzed using linear discriminant analysis. The dependent variable which is the genotype group is classified into CB57L/6J (0) and RD10 (1) while the predictor or independent variable is five numbers Summary of area and eccentricity which made up of Ten independent variables. The Linear Discriminant analysis of the above data set is done assumed equal variance. The pooled covariance matrix is -32.29224 while the error rate is .00167.

**Table 5.1 Number of Observations and Percent Classified into genotype using Linear Determinant Analysis (five summary statistics)**

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>100 | 0<br>0 | 50<br>100 |
| 1 | 2<br>3.33 | 58<br>96.67 | 60<br>100 |
| Total | 52<br>47.27 | 58<br>52.73 | 110<br>100 |
| Priors | 0.5 | 0.5 | |

From Table 5.1 above we see 110 individual eyes 100 percent Genotype CB57L/6J (0) were correctly classified as better than 96.67 percent Genotype RD10 Classified correctly. Overall, 1.67 percent of the individual Genotype of each eye is misclassified.

### 5.2 Quadratic Discriminant Analysis of Genotype with Statistics Morphometric measure of each eye.

The five summary statistics of 110 individual eyes are analyzed. The dependent variable which is the group is classified into CB57L/6J and RD10 while the predictor or independent variable are five number summary of area and eccentricity which made up of Ten independent variables. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance Matrix for CB57L/6J and RD10 are -35.07331 and -40.12572 respectively while the error rate 0.00.

**Table 5.2 Number of Observations and Percent Classified into genotype using Quadratic Determinant Analysis (five summary statistics)**

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| **0** | 50<br>100 | 0<br>0 | 50<br>100 |
| **1** | 0<br>0 | 60<br>100 | 60<br>100 |
| **Total** | 50<br>45.45 | 60<br>54.55 | 110<br>100 |
| **Priors** | 0.5 | 0.5 | |

From Table 5.2 above we see 110 individual eyes are 100 percent correctly classified into Genotype CB57L/6J (0) and Genotype RD10 with 0% overall error rate.

### 5.3 Linear Discriminant Analysis of Age Group with Statistics Morphometric measure of Cells for each eyes

The five summary statistics of 110 individual eye which will be analyzed using linear discriminant analysis.The dependent variable which is the group is classified into young (0) and old (1) while the predictor or independent variable are five number summary of areashape and eccentricity which made up of Ten independent variables. The Linear Discriminant analysis of the above data set is done assumed equal variance. The pooled covariance Matrix is -31.61814 while the error rate is 0.0999.

**Table 5.3 Number of Observations and Percent Classified into Age Group using Linear Determinant Analysis (five summary statistics)**

| Age Group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 49 | 5 | 54 |
| | 90.74 | 9.26 | 100 |
| 1 | 6 | 50 | 56 |
| | 10.71 | 89.29 | 100 |
| Total | 55 | 55 | 110 |
| | 50 | 50 | 100 |
| Priors | 0.5 | 0.5 | |

From Table 5.3 above we see 110 individual eyes of which 90.74 percent young group (0) classified correctly which is better than 89.29 percent correctly classified as older group(1). Overall, 9.99% of the observations were mis-classified.

**5.4 Quadratic Discriminant Analysis of Age group with Statistics Morphometric measure of cell for each eyes.**

The five summary statistics of 110 individual eyes are analyzed. The dependent variable which is the group is classified into young (0) and old (1) while the predictor or independent variable are five number summary of areashape and eccentricity which made up of Ten independent variables. The Quadratic Discriminant analysis of the above data set is done assumed unequal variance. The pooled covariance Matrix for are young (0) and old (1) -43.58733 and -30.32510 respectively while the error rate 0.0724.

**Table 5.4 Number of Observations and Percent Classified into Age group using Quadratic Determinant Analysis (five summary statistics)**

| Age group | 0 | 1 | Total |
|---|---|---|---|
| 0 | 51<br>94.44 | 3<br>5.56 | 54<br>100 |
| 1 | 5<br>8.93 | 51<br>91.07 | 56<br>100 |
| Total | 56<br>50.91 | 54<br>49.09 | 110<br>100 |
| Priors | 0.5 | 0.5 | |

From Table 5.4 above we see 110 individual eyes of which 90.44 percent young group (0) classified correctly which is not better than 91.07 percent correctly classified as older group(1). Overall, 7.24% of the observations were mis-classified.

**Table 5.5 Group Classification for Genotype Group (five summary statistics)**

| Class Level Information | | | | | |
|---|---|---|---|---|---|
| **genotype1** | **Variable Name** | **Frequency** | **Weight** | **Proportion** | **Prior Probability** |
| **0** | _0 | 50 | 50.0000 | 0.454545 | 0.500000 |
| **1** | _1 | 60 | 60.0000 | 0.545455 | 0.500000 |

Table 4.5 shows that the RD10 (1) contributes most to Genotype group separations which has 54.8 percent of 110 Eyes.

**Table 5.6 Group Classification for Genotype Age group (five summary statistics)**

| Class Level Information | | | | | |
|---|---|---|---|---|---|
| **age1** | **Variable Name** | **Frequency** | **Weight** | **Proportion** | **Prior Probability** |
| **0** | _0 | 54 | 54.0000 | 0.490909 | 0.500000 |
| **1** | _1 | 56 | 56.0000 | 0.509091 | 0.500000 |

Table 5.5 shows that the older group (0) contributes most to age group separations which has 50.9 percent of 110 Eyes.

# 6 CONCLUSIONS

Using linear discriminant analysis, From Table 3.1 above we see 199804 cells observation of which 45.31 percent were correctly misclassified as Genotype CB57L/6J (0) while than 36.94 percent Genotype RD10(1) is misclassified. Although using quadratic discriminant analysis, the misclassification rate for Genotype CB57L/6J (0), Genotype RD10 (1) increase to 55.93 percent and 62.56 percent respectively. The result is not good.

Using Linear discriminant analysis for age group, From Table 3.3 above we see 199804 cells observation of which 40.28 percent were correctly misclassified young group (0) which is 50.52 percent correctly misclassified as older group(1). From Table 3.4 above, using quadratic discriminant analysis, the misclassification rate for percent young group (0 , older group(1) ) reduce to 19.50 percent and increase to 69.94 percent respectively. The result is not balanced and good.

This accuracy is disappointingly low, because we know from previous fPCA [jiang et al 2012], the classification accuracy for four group (young C57BL/6J, old C57BL/6J, young rd10 and old rd10) can reach as high as 98%.

To improve the result, individual Eyes and Genotype with Morphometric cells measures is broken into 5 percentile Area , 50 percentile Area , 25$^{th}$ percentile Area , 75 percentile Area , 95 percentile Area , 5 percentile Eccentricity, 25 percentile Eccentricity, 50 percentile Eccentricity, 75 percentile Eccentricity and 95 percentile Eccentricity measure. Overall, using Linear and quadratic discriminant analysis the misclassification for the genotypes, age group reduces to zero (0). Thus variables of each eye provide a better classification as overall error rate reduce practical to zero in case of genotype and 9.9 percent in case of age.

What we learned is that the morphometric measures of individual cells do not offer good classification for eye's genotype and age; individual cells from all eyes mixed together is equivalent to assume all cells are independent, which is missing their important correlations within the eyes. It is the eye level infor-

mation, through dimensional reduction methods on the many thousands cells in the eye, that will offer best classification of the eye's genotype and age.

Based on the Linear discriminant analysis and Quadratic discriminant analysis the combination of Area and shape provide a better classification.

Also, Linear discriminant analysis and Quadratic discriminant analysis will discovered that Genotype provide a good classification result than Age.

Lastly, Quadratic discriminant analysis (QDA) provides a better result than linear discriminant analysis (LDA) but both analyses in some cases produce the same result.

**REFERENCES**


1. Bedient CE, Magrina JF, Noble BN, et al. Comparison of robotic and laparoscopic myomectomy. Am J Obstet Gynecol 2009; 201
2. Richard A. Johnson , Dean W. Wichern Applied Multivariate Statistical Analysis. 6th Edition 2007.
3. Bernard Flury A First Course In Multivarate Statistics 1st Edition 1997.
4. Willian R Dillon, Mathew Goldstein Multivariate Analysis 1st Edition 1941
5. Barbara G. Tabachnick, Linda S. Fidell Using Multivariate Statistics 4th Edition 2001
6. Sanjay Matange, Dan Heath Statistical Graphics Procedure by Example 1st Edition 2011
7. John Shawe-Taylor, Nello Cristianini Kernel Methods For Pattern Analysis 1st Edition 2004.
8. Lora D. Delwiche and Susan J. Slaughter The Little Sas Book 4th Edition 2008.
9. Micah A. Chrenek, Nupur Dalal, Christopher Gardner, Hans GrossniklausGrossniklaus and Yi Jiang, et al, Analysis of the RPE Sheet in the rd10 Retinal Degeneration Model, *Advances in Experimental Medicine and Biology*, 1, Volume 723, Retinal Degenerative Diseases, Part 8, Pages 641-647, 2012.
10. Klein R, Klein BE, Linton KL Prevalence of age-related maculopathy, The Beaver Dam Eye Study,*Ophthalmology* 99: 933–943, 1992.
11. Klein R, Klein BE, Jensen SC, Meuer SM, The five-year incidence and progression of age-related maculopathy: the Beaver Dam Eye Study. *Ophthalmology* 104: 7–21, 1997.

**APPENDICES**

Appendix A : SAS CODE FOR Data Cleaning

```
proc import datafile="f:\rd10_723_1.xls" out=oney dbms=excel;
run;
data wa1;
set oney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="f:\rd10_723_2.xls" out=twoy dbms=excel;
run;
data wa2;
set twoy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="f:\rd10_330_3.xls" out=threey dbms=excel;
run;
data wa3;
set threey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="f:\rd10_330_4.xls" out=foury dbms=excel;
run;
data wa4;
set foury (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="f:\rd10_330_5.xls" out=fivey dbms=excel;
run;
data wa5;
set fivey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_30_6.xls" out=sixy dbms=excel;
run;
data wa6;
set sixy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_30_7.xls" out=seveny dbms=excel;
run;
data wa7;
set seveny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_30_8.xls" out=eighty dbms=excel;
run;
data wa8;
set eighty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
```

```
run;

proc import datafile="F:\rd10_61_9.xls" out=niney dbms=excel;
run;
data wa9;
set niney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_61_10.xls" out=teny dbms=excel;
run;
data wa10;
set teny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_11.xls" out=eleveny dbms=excel;
run;
data wa11;
set eleveny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_12.xls" out=twevey dbms=excel;
run;
data wa12;
set twevey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_100_13.xls" out=thirteeny dbms=excel;
run;
data wa13;
set thirteeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_14.xls" out=forteeny dbms=excel;
run;
data wa14;
set forteeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_15.xls" out=fifteeny dbms=excel;
run;
data wa15;
set fifteeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_16.xls" out=sixteeny dbms=excel;
run;
data wa16;
set sixteeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_17.xls" out=seventeeny dbms=excel;
run;
data wa17;
set seventeeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run
```

```
proc import datafile="F:\rd10_100_18.xls" out=eighteeny dbms=excel;
run;
data wa18;
set eighteeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_19.xls" out=nineteeny dbms=excel;
run;
data wa19;
set nineteeny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_60_20.xls" out=twentyy dbms=excel;
run;
data wa20;
set twentyy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_60_21.xls" out=twentyoney dbms=excel;
run;
data wa21;
set twentyoney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_60_22.xls" out=twentytwoy dbms=excel;
run;
data wa22;
set twentytwoy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_60_23.xls" out=twentythreey dbms=excel;
run;
data wa23;
set twentythreey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_60_24.xls" out=twentyfoury dbms=excel;
run;
data wa24;
set twentyfoury (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_60_25.xls" out=twentyfivey dbms=excel;
run;
data wa25;
set twentyfivey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_26.xls" out=twentysixy dbms=excel;
run;
data wa26;
set twentysixy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_27.xls" out=twentyseveny dbms=excel;
```

```
run;
data wa27;
set twentyseveny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_61_28.xls" out=twentyeighty dbms=excel;
run;
data wa28;
set twentyeighty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_61_29.xls" out=twentyniney dbms=excel;
run;
data wa29;
set twentyniney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_61_30.xls" out=thirtyy dbms=excel;
run;
data wa30;
set thirtyy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_45_31.xls" out=thirtyoney dbms=excel;
run;
data wa31;
set thirtyoney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_32.xls" out=thirtytwoy dbms=excel;
run;
data wa32;
set thirtytwoy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_45_33.xls" out=thirtythreey dbms=excel;
run;
data wa33;
set thirtythreey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_45_34.xls" out=thirtyfoury dbms=excel;
run;
data wa34;
set thirtyfoury (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_35.xls" out=thirtyfivey dbms=excel;
run;
data wa35;
set thirtyfivey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_36.xls" out=thirtysixy dbms=excel;
```

```
run;
data wa36;
set thirtysixy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_37.xls" out=thirtyseveny dbms=excel;
run;
data wa37;
set thirtyseveny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_45_38.xls" out=thirtyeighty dbms=excel;
run;
data wa38;
set thirtyeighty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_45_39.xls" out=thirtyniney dbms=excel;
run;
data wa39;
set thirtyniney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_40.xls" out=fortyy dbms=excel;
run;
data wa40;
set fortyy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_45_41.xls" out=fortyoney dbms=excel;
run;
data wa41;
set fortyoney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_732_42.xls" out=fortytwoy dbms=excel;
run;
data wa42;
set fortytwoy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_732_43.xls" out=fortythreey dbms=excel;
run;
data wa43;
set fortythreey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_732_44.xls" out=fortyfoury dbms=excel;
run;
data wa44;
```

```
set  fortyfoury (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_45.xls" out=fortyfivey dbms=excel;
run;
data wa45;
set  fortyfivey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_46.xls" out=fortysixy dbms=excel;
run;
data wa46;
set  fortysixy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_47.xls" out=fortyseveny dbms=excel;
run;
data wa47;
set  fortyseveny (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_48.xls" out=fortyeighty dbms=excel;
run;
data wa48;
set  fortyeighty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_49.xls" out=fortyniney dbms=excel;
run;
data wa49;
set  fortyniney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_50.xls" out=fiftyy dbms=excel;
run;
data wa50;
set  fiftyy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_51.xls" out=fiftyoney dbms=excel;
run;
data wa51;
set fiftyoney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_52.xls" out= fiftytwoy dbms=excel;
run;
data wa52;
```

```
set  fiftytwoy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_180_53.xls" out=fiftythreey dbms=excel;
run;
data wa53;
set  fiftythreey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_180_54.xls" out=fiftyfoury dbms=excel;
run;
data wa54;
set  fiftyfoury (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_180_55.xls" out=fiftyfivey dbms=excel;
run;
data wa55;
set  fortyfivey (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_180_56.xls" out= fiftysixy dbms=excel;
run;
data wa56;
set  fiftysixy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\rd10_180_57.xls" out= fiftyseveny dbms=excel;
run;
data wa57;
set  fiftysixy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_58.xls" out=fiftyeighty dbms=excel;
run;
data wa58;
set  fiftyeighty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_59.xls" out=fiftyniney dbms=excel;
run;
data wa59;
set  fiftyniney (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\rd10_100_60.xls" out= sixtyy dbms=excel;
run;
data wa60;
set  sixtyy (keep=Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_p720_34.xls" out=one dbms=excel;
run;
data ay1;
set one (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
```

```
run;


proc import datafile="F:\C57BL6J_p720_33.xls" out=two dbms=excel;
run;
data ay2;
set two (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_p720_32.xls" out=three dbms=excel;
run;
data ay3;
set three (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_p30_100.xls" out=four dbms=excel;
run;
data ay4;
set four (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_p30_99.xls" out=five dbms=excel;
run;
data ay5;
set five (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_p30_98.xls" out=six dbms=excel;
run;
data ay6;
set six (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_p30_96.xls" out=seven dbms=excel;
run;
data ay7;
set seven (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_330_15.xls" out=eight dbms=excel;
run;
data by1;
set  eight (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_330_16.xls" out=nine dbms=excel;
run;
data by2;
set  nine (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_330_17.xls" out=ten dbms=excel;
run;
data by3;
```

```
set  ten (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_60_18.xls" out=eleven dbms=excel;
run;
data by4;
set eleven (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_60_19.xls" out=tweve dbms=excel;
run;
data by5;
set  tweve (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_60_20.xls" out=thirteen dbms=excel;
run;
data by6;
set  thirteen (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_60_21.xls" out=fourteen dbms=excel;
run;
data by7;
set  fourteen (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_30_22.xls" out=fifteen dbms=excel;
run;
data by8;
set  fifteen (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_30_23.xls" out=thirtyseven dbms=excel;
run;
data by9;
set  thirtyseven (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_30_24.xls" out=thirtyeight dbms=excel;
run;
data by10;
set   thirtyeight (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_30_25.xls" out=thirtynine dbms=excel;
run;
data by11;
set thirtynine (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_30_26.xls" out=forty dbms=excel;
run;
data by12;
set  forty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run
```

```
proc import datafile="F:\C57BL6J_30_27.xls" out=fortyone dbms=excel;
run;
data by13;
set  fortyone (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_61_28.xls" out=fortytwo dbms=excel;
run;
data by14;
set  fortytwo (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_61_29.xls" out=fortythree dbms=excel;
run;
data by15;
set  fortythree (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_61_30.xls" out=fortyfour dbms=excel;
run;
data by16;
set  fortyfour (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_61_31.xls" out=fortyfive dbms=excel;
run;

data by17;
set  fortyfive (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_61_32.xls" out=fortysix dbms=excel;
run;
data by18;
set  fortysix (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_45_33.xls" out=fortyseven dbms=excel;
run;
data by19;
set  fortyseven (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_45_34.xls" out=fortyeight dbms=excel;
run;
data by20;
set  fortyeight (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_45_35.xls" out=fortynine dbms=excel;
run;
data by21;
set  fortynine (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
```

```
run;
proc import datafile="F:\C57BL6J_45_36.xls" out=fifty dbms=excel;
run;

data by22;
set  fifty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_45_37.xls" out=fiftyone dbms=excel;
run;
data by23;
set  fiftyone (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_45_38.xls" out=fiftytwo dbms=excel;
run;
data by24;
set  fiftytwo (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_45_39.xls" out=fiftythree dbms=excel;
run;
data by25;
set  fiftythree (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_45_40.xls" out=fiftyfour dbms=excel;
run;
data by26;
set  fiftyfour (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_720_41.xls" out=fiftyfive dbms=excel;
run;

data by27;
set  fiftyfive (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_720_42.xls" out=fiftysix dbms=excel;
run;
data by28;
set  fiftysix (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_720_43.xls" out=fiftyseven dbms=excel;
run;
data by29;
set fiftyseven (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_180_44.xls" out=fiftyeight dbms=excel;
```

```
run;
data by30;
set  fiftyeight (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_180_45.xls" out=fiftynine dbms=excel;
run;
data by31;
set  fiftynine (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_180_46.xls" out=sixty dbms=excel;
run;
data by32;
set  sixty (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_330_49.xls" out=sixtythree dbms=excel;
run;
data by33;
set  sixtythree (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_722_53.xls" out=sixtyseven dbms=excel;
run;
data by34;
set  sixtyseven (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_722_54.xls" out=sixtyeight dbms=excel;
run;
data by35;
set  sixtyeight (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_722_55.xls" out=sixtynine dbms=excel;
run;
data by36;
set  sixtynine (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_722_56.xls" out=seventy dbms=excel;
run;
data by37;
set  seventy (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

proc import datafile="F:\C57BL6J_722_57.xls" out=seventyone dbms=excel;
run;
data by38;
set  seventyone (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
```

```
proc import datafile="F:\C57BL6J_180_58.xls" out=seventytwo dbms=excel;
run;
data by39;
set seventytwo (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_180_59.xls" out=seventythree dbms=excel;
run;
data by40;
set seventythree (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_180_60.xls" out=seventyfour dbms=excel;
run;
data by41;
set seventyfour (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;
proc import datafile="F:\C57BL6J_180_61.xls" out=seventyfive dbms=excel;
run;
data by42;
set  seventyfive (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
rum;
proc import datafile="F:\C57BL6J_180_62.xls" out=seventysix dbms=excel;
run;
data by43;
set  seventysix (keep =Age Genotype AreaShape_Area AreaShape_Eccentricity) ;
run;

data win;
merge ay1 ay2 ay3 ay4 ay5 ay6 ay7 wa1 wa2 wa3 wa4 wa5 wa6 wa7 wa8 wa9
wa10 wa11 wa12 wa13 wa14 wa15 wa16 wa17 wa18 wa19 wa20 wa21 wa22
wa23 wa24 wa25 wa26 wa27 wa28 wa29 wa30 wa31 wa32 wa33 wa34 wa35
wa36 wa37 wa38 wa39 wa40 wa41 wa42 wa43 wa44 wa45 wa46 wa47 wa48
wa49 wa50 wa51 wa52 wa53 wa54 wa55 wa56 wa57 wa58 wa59 wa60 ay1 ay2 ay3 ay4 ay5 ay6 ay7
by1 by2 by3 by4 by5 by6 by7 by8 by9 by10 by11 by12 by13 by14 by15 by16
by17 by18 by19 by20 by21 by22 by23 by24 by25 by26 by27 by28 by29
by30 by31 by32 by33 by34 by35 by36 by37 by38 by39 by40 by41 by42 by43;
by genotype age;run;

data analy1;
set win;
genotype1=0;
if genotype="c57BL/6J" then genotype1=0;
if genotype="C57BL/6J" then genotype1=0;
if genotype="rd10" then genotype1=1;
 age1=0;
if age<70 then age1=0;
if age=>70 then age1=1;
run;
```

```
proc sort data=analy1; by genotype1;run;

*****Density Curve****;

proc import datafile="F:\CAT.xlsx" out=housy dbms=excel;
run;
proc print data=housy;run;
proc sort data=housy; by genotype;run;


data win;
SET housy;
          genotype1=0;
id=_n_;
if genotype= 'C57BL6J'  then genotype1=0;
if genotype= 'RD10'  then genotype1=1;
 age1=0;
if age<70 then age1=0;
if age=>70 then age1=1;
age2=0;
if age<320 then age2=0;
if age=>320 then age2=1;
age3=0;
if age<180 then age3=0;
if age=>180 then age3=1;
run;
proc print data=win;run;


data shola1 shola2;
set win;
if genotype1=0 then output shola1;
else output shola2;
run;
```

Appendix B : SAS CODE FOR Density Curve Of Genotype C57BL6J

```
*****'C57BL6J'****;

*****age70****;

ods output kolsmir2stats=age_area_ks_stats;
ods select wilcoxontest kolsmir2stats;
proc npar1way data=shola1 wilcoxon edf;
class age1;
var AreaShape_Area;
run;
ods select all;

proc sort data=shola1;
by age1;
run;
ods select none;

proc kde data=shola1;
by age1;
univar AreaShape_Area/ out=kdeout;
run;
ods select all;
proc print data=age_area_ks_stats;run;

data _null_;
set age_area_ks_stats;
if label2 eq 'D' then call symput('dvalue', substr(cvalue2, 1, 5));
if label2 eq 'pr >KSa' then call symput('pvalue', substr(cvalue2, 1, 4));
run;

symbol1 i=j w=5 l=1 v=none c=black;
symbol2 i=j w=5 l=2 v=none c=blue;
title 'Density Curve of Area for Age 70 and Genotype C57BL/6J';
pattern1 color=grayBB;
proc gplot data=kdeout;
plot density*value= age1/legend  haxis=125 to 180 by 5;
run; quit;

*****age180****;

ods output kolsmir2stats=age_area_ks_stats;
ods select wilcoxontest kolsmir2stats;
proc npar1way data=shola1 wilcoxon edf;
class age3;
var AreaShape_Area;
run;
ods select all;

proc sort data=shola1;
by age3;
run;
ods select none;

proc kde data=shola1;
```

```
by age3;
univar AreaShape_Area/ out=kdeout1;
run;
ods select all;
proc print data=age_area_ks_stats;run;

data _null_;
set age_area_ks_stats;
if label2 eq 'D' then call symput('dvalue', substr(cvalue2, 1, 5));
if label2 eq 'pr >KSa' then call symput('pvalue', substr(cvalue2, 1, 4));
run;

symbol1 i=j w=5 l=1 v=none c=black;
symbol2 i=j w=5 l=2 v=none c=blue;
title 'Density Curve of Area for Age 180 and Genotype C57BL/6J ';
pattern1 color=grayBB;
proc gplot data=kdeout1;
plot density*value= age3/legend  haxis=125 to 180 by 5;
run; quit;

*****age320****;

ods output kolsmir2stats=age_area_ks_stats;
ods select wilcoxontest kolsmir2stats;
proc npar1way data=shola1 wilcoxon edf;
class age2;
var AreaShape_Area;
run;
ods select all;

proc sort data=shola1;
by age2;
run;
ods select none;

proc kde data=shola1;
by age2;
univar AreaShape_Area/ out=kdeout2;
run;
ods select all;
proc print data=age_area_ks_stats;run;

data _null_;
set age_area_ks_stats;
if label2 eq 'D' then call symput('dvalue', substr(cvalue2, 1, 5));
if label2 eq 'pr >KSa' then call symput('pvalue', substr(cvalue2, 1, 4));
run;

symbol1 i=j w=5 l=1 v=none c=black;
symbol2 i=j w=5 l=2 v=none c=blue;
title 'Density Curve of Area for Age 320 and Genotype C57BL/6J ';
pattern1 color=grayBB;
proc gplot data=kdeout2;
plot density*value= age2/legend  haxis=125 to 180 by 5;
run; quit;
```

Appendix C : SAS CODE FOR Density Curve OF Genotype RD10

```
*****age70 RD10****;

ods output kolsmir2stats=age_area_ks_stats;
ods select wilcoxontest kolsmir2stats;
proc npar1way data=shola1 wilcoxon edf;
class age1;
var AreaShape_Area;
run;
ods select all;

proc sort data=shola2;
by age1;
run;
ods select none;

proc kde data=shola2;
by age1;
univar AreaShape_Area/ out=kdeout;
run;
ods select all;
proc print data=age_area_ks_stats;run;

data _null_;
set age_area_ks_stats;
if label2 eq 'D' then call symput('dvalue', substr(cvalue2, 1, 5));
if label2 eq 'pr >KSa' then call symput('pvalue', substr(cvalue2, 1, 4));
run;

symbol1 i=j w=5 l=1 v=none c=black;
symbol2 i=j w=5 l=2 v=none c=blue;
title 'Density Curve of Area for Age 70 and Genotype Rd10';
pattern1 color=grayBB;
proc gplot data=kdeout;
plot density*value= age1/legend  haxis=135 to 190 by 10;
run; quit;

*****age180****;

ods output kolsmir2stats=age_area_ks_stats;
ods select wilcoxontest kolsmir2stats;
proc npar1way data=shola2 wilcoxon edf;
class age3;
var AreaShape_Area;
run;
ods select all;

proc sort data=shola2;
by age3;
run;
ods select none;

proc kde data=shola2;
by age3;
univar AreaShape_Area/ out=kdeout1;
run;
```

```
ods select all;
proc print data=age_area_ks_stats;run;

data _null_;
set age_area_ks_stats;
if label2 eq 'D' then call symput('dvalue', substr(cvalue2, 1, 5));
if label2 eq 'pr >KSa' then call symput('pvalue', substr(cvalue2, 1, 4));
run;

symbol1 i=j w=5 l=1 v=none c=black;
symbol2 i=j w=5 l=2 v=none c=blue;
title 'Density Curve of Area for Age 180 and Genotype C57BL/6J ';
pattern1 color=grayBB;
proc gplot data=kdeout1;
plot density*value= age3/legend  haxis=135 to 185 by 5;
run; quit;




*****age320****;

ods output kolsmir2stats=age_area_ks_stats;
ods select wilcoxontest kolsmir2stats;
proc npar1way data=shola2 wilcoxon edf;
class age2;
var AreaShape_Area;
run;
ods select all;

proc sort data=shola2;
by age2;
run;
ods select none;

proc kde data=shola2;
by age2;
univar AreaShape_Area/ out=kdeout2;
run;
ods select all;
proc print data=age_area_ks_stats;run;

data _null_;
set age_area_ks_stats;
if label2 eq 'D' then call symput('dvalue', substr(cvalue2, 1, 5));
if label2 eq 'pr >KSa' then call symput('pvalue', substr(cvalue2, 1, 4));
run;

symbol1 i=j w=5 l=1 v=none c=black;
symbol2 i=j w=5 l=2 v=none c=blue;
title 'Density Curve of Area for Age 320 and Genotype C57BL/6J ';
pattern1 color=grayBB;
proc gplot data=kdeout2;
plot density*value= age2/legend  haxis=135 to 185 by 5;
run; quit;
```

**Appendix D: SAS CODE For Linear/Quadratic Discriminant Function of Cells measure**

```
********Genotype*********;

ods html;
proc discrim data=analy1  outstat=agestat pool=yes  crossvalidate;
      class genotype1;
      priors Equal;
      var AreaShape_Area AreaShape_Eccentricity;
      title2 'Using Linear Discriminant Function';
   run;
ods html close;

ods html;
proc discrim data=analy1  outstat=agestat pool=no  crossvalidate;
      class genotype1;
      priors Equal;
      var AreaShape_Area AreaShape_Eccentricity;
      title2 'Using Quadratic Discriminant Function Genotype';
   run;
ods html close;


********Agegroup*********;

ods html;
proc discrim data=analy1  outstat=agestat pool=yes  crossvalidate;
      class age1;
      priors Equal;
      var AreaShape_Area AreaShape_Eccentricity;
      title2 'Using Linear Discriminant Function';
   run;
ods html close;

ods html;
proc discrim data=analy1  outstat=agestat pool=no  crossvalidate;
      class age1;
      priors Equal;
      var AreaShape_Area AreaShape_Eccentricity;
      title2 'Using Quadratic Discriminant Function';
   run;
ods html close;
```

**Appendix E: SAS CODE FOR Linear/Quadratic Discriminant Function of Statistics measure of Cell**

```
*******Summary Statistics dataset*******;


proc import datafile="F:\CAT.xlsx" out=housy dbms=excel;
run;
proc print data=housy;run;
proc sort data=housy; by genotype;run;


data win;
SET housy;
      genotype1=0;
id=_n_;
if genotype= 'C57BL6J'  then genotype1=0;
if genotype= 'RD10'  then genotype1=1;
 age1=0;
if age<70 then age1=0;
if age=>70 then age1=1;
run;
proc print data=win;run;



*****'Using Linear Discriminant Function'*****;


ods html;
proc discrim data=win  outstat=agestat pool=yes  crossvalidate;
      class genotype1;
      priors Equal;
      var  AreaShape_Area AreaShape_Eccentricity Areafift  Areatwentyfift
Areasevenfive   Areaninetyfive EccentFive
Eccenttwentyfive Eccentseventyfive eccentninetyfive;
      title2 'Using Linear Discriminant Function';
   run;
ods html close;


ods html;
proc discrim data=win  outstat=agestat pool=no  crossvalidate;
      class genotype1;
      priors Equal;
      var  AreaShape_Area AreaShape_Eccentricity Areafift  Areatwentyfift
Areasevenfive   Areaninetyfive EccentFive
Eccenttwentyfive Eccentseventyfive eccentninetyfive;
      title2 'Using Quadratic Discriminant Function Genotype';
   run;
ods html close;



***'Age'****;
```

```
ods html;
proc discrim data=win  outstat=agestat pool=yes  crossvalidate;
      class age1;
      priors Equal;
      var AreaShape_Area AreaShape_Eccentricity Areafift  Areatwentyfift
Areasevenfive   Areaninetyfive EccentFive
Eccenttwentyfive Eccentseventyfive eccentninetyfive;
      title2 'Using Linear Discriminant Function';
   run;
ods html close;
```

```
ods html;
proc discrim data=win  outstat=agestat pool=no  crossvalidate;
      class age1;
      priors Equal;
      var AreaShape_Area AreaShape_Eccentricity Areafift  Areatwentyfift
Areasevenfive   Areaninetyfive EccentFive
Eccenttwentyfive Eccentseventyfive eccentninetyfive ;
      title2 'Using Quadratic Discriminant Function';
   run;
ods html close;
```

## Appendix F : SAS Output for Using Linear Discriminant Function Genotype

```
                                The SAS System                    09:3
                        Using Linear Discriminant Function

                              The DISCRIM Procedure

          Total Sample Size   199804          DF Total            199803
          Variables                2          DF Within Classes   199802
          Classes                  2          DF Between Classes       1


                    Number of Observations Read         199804
                    Number of Observations Used         199804


                              Class Level Information

                 Variable                                                Prior
     genotype1   Name        Frequency       Weight    Proportion    Probability

             0   _0             117659       117659      0.588872       0.500000
             1   _1              82145        82145      0.411128       0.500000


                         Pooled Covariance Matrix Information

                                        Natural Log of the
                          Covariance    Determinant of the
                         Matrix Rank     Covariance Matrix

                                   2               4.52443
```

The SAS System 09:34 Thursday, June 14, 2012 11853

Using Linear Discriminant Function

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} \, (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to genotype1

| From genotype1 | 0 | 1 |
|---|---|---|
| 0 | 0 | 0.19089 |
| 1 | 0.19089 | 0 |

Linear Discriminant Function

$$\text{Constant} = -.5 \, \bar{X}_j' \, COV^{-1} \, \bar{X}_j \qquad \text{Coefficient Vector} = COV^{-1} \, \bar{X}_j$$

Linear Discriminant Function for genotype1

| Variable | Label | 0 | 1 |
|---|---|---|---|
| Constant | | -12.18433 | -13.80731 |
| AreaShape_Area | AreaShape_Area | 0.03895 | 0.03720 |
| AreaShape_Eccentricity | AreaShape_Eccentricity | 27.64930 | 30.47758 |

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 64348 | 53311 | 117659 |
| | 54.69 | 45.31 | 100.00 |
| 1 | 30344 | 51801 | 82145 |
| | 36.94 | 63.06 | 100.00 |
| Total | 94692 | 105112 | 199804 |
| | 47.39 | 52.61 | 100.00 |
| Priors | 0.5 | 0.5 | |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Linear Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.4531 | 0.3694 | 0.4112 |
| Priors | 0.5000 | 0.5000 | |

The SAS System 09:34 Thursday, June 14, 2012 11856

Using Linear Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 64347 | 53312 | 117659 |
| | 54.69 | 45.31 | 100.00 |
| 1 | 30345 | 51800 | 82145 |
| | 36.94 | 63.06 | 100.00 |
| Total | 94692 | 105112 | 199804 |
| | 47.39 | 52.61 | 100.00 |
| Priors | 0.5 | 0.5 | |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Cross-validation Summary using Linear Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.4531 | 0.3694 | 0.411 |
| Priors | 0.5000 | 0.5000 | |

## Appendix G: SAS output Using Quadratic Discriminant Function Genotype

The SAS System 09:34 Thursday, June 14, 2012 11858
Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure

| | | | |
|---|---|---|---|
| Total Sample Size | 199804 | DF Total | 199803 |
| Variables | 2 | DF Within Classes | 199802 |

| | | | |
|---|---|---|---|
| Classes | 2 | DF Between Classes | 1 |

| | |
|---|---|
| Number of Observations Read | 199804 |
| Number of Observations Used | 199804 |

Class Level Information

| genotype1 | Variable Name | Frequency | Weight | Proportion | Prior Probability |
|---|---|---|---|---|---|
| 0 | _0 | 117659 | 117659 | 0.588872 | 0.500000 |
| 1 | _1 | 82145 | 82145 | 0.411128 | 0.500000 |

Within Covariance Matrix Information

| genotype1 | Covariance Matrix Rank | Natural Log of the Determinant of the Covariance Matrix |
|---|---|---|
| 0 | 2 | 4.47985 |
| 1 | 2 | 4.58455 |

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV_j^{-1} \, (\bar{X}_i - \bar{X}_j) + \ln |COV_j|$$

Generalized Squared Distance to genotype1

| From genotype1 | 0 | 1 |
|---|---|---|
| 0 | 4.47985 | 4.77469 |
| 1 | 4.67140 | 4.58455 |

The SAS System 09:34 Thursday, June 14, 2012 11860

Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j|$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 65809 | 51850 | 117659 |
| | 55.93 | 44.07 | 100.00 |
| 1 | 30756 | 51389 | 82145 |
| | 37.44 | 62.56 | 100.00 |
| Total | 96565 | 103239 | 199804 |
| | 48.33 | 51.67 | 100.00 |
| Priors | 0.5 | 0.5 | |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Quadratic Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.4407 | 0.3744 | 0.4075 |
| Priors | 0.5000 | 0.5000 | |

The SAS System 09:34 Thursday, June 14, 2012 11862

Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Cross-validation Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)j}^{-1} \, (X-\bar{X}_{(X)j}) + \ln |COV_{(X)j}|$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 65808 | 51851 | 117659 |
| | 55.93 | 44.07 | 100.00 |
| 1 | 30757 | 51388 | 82145 |
| | 37.44 | 62.56 | 100.00 |
| Total | 96565 | 103239 | 199804 |
| | 48.33 | 51.67 | 100.00 |
| Priors | 0.5 | 0.5 | |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Cross-validation Summary using Quadratic Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.4407 | 0.3744 | 0.4076 |
| Priors | 0.5000 | 0.5000 | |

## Appendix H: SAS Output for Using Linear Discriminant Function on Age Group

Using Linear Discriminant Function

The DISCRIM Procedure

| | | | |
|---|---|---|---|
| Total Sample Size | 199804 | DF Total | 199803 |
| Variables | 2 | DF Within Classes | 199802 |
| Classes | 2 | DF Between Classes | 1 |

| | |
|---|---|
| Number of Observations Read | 199804 |
| Number of Observations Used | 199804 |

Class Level Information

| age1 | Variable Name | Frequency | Weight | Proportion | Prior Probability |
|---|---|---|---|---|---|
| 0 | _0 | 115580 | 115580 | 0.578467 | 0.500000 |
| 1 | _1 | 84224 | 84224 | 0.421533 | 0.500000 |

Pooled Covariance Matrix Information

| Covariance Matrix Rank | Natural Log of the Determinant of the Covariance Matrix |
|---|---|
| 2 | 4.55498 |

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} \, (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to age1

| From age1 | 0 | 1 |
|---|---|---|
| 0 | 0 | 0.06046 |
| 1 | 0.06046 | 0 |

Linear Discriminant Function

$$\text{Constant} = -.5 \, \bar{X}_j' \, COV^{-1} \, \bar{X}_j \qquad \text{Coefficient Vector} = COV^{-1} \, \bar{X}_j$$

Linear Discriminant Function for age1

| Variable | Label | 0 | 1 |
|---|---|---|---|
| Constant | | -12.21612 | -13.32715 |
| AreaShape_Area | AreaShape_Area | 0.03838 | 0.04162 |
| AreaShape_Eccentricity | AreaShape_Eccentricity | 27.62889 | 28.47721 |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 69023 | 46557 | 115580 |
| | 59.72 | 40.28 | 100.00 |
| 1 | 42551 | 41673 | 84224 |
| | 50.52 | 49.48 | 100.00 |
| Total | 111574 | 88230 | 199804 |
| | 55.84 | 44.16 | 100.00 |
| Priors | 0.5 | 0.5 | |

The SAS System 09:34 Thursday, June 14, 2012 11873

Using Linear Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Linear Discriminant Function

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.4028 | 0.5052 | 0.4540 |
| Priors | 0.5000 | 0.5000 | |

Classification Summary for Calibration Data: WORK.ANALY1
Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = exp(-.5\, D_j^2(X)) \,/\, SUM_k \, exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 69023 | 46557 | 115580 |
| | 59.72 | 40.28 | 100.00 |
| 1 | 42552 | 41672 | 84224 |
| | 50.52 | 49.48 | 100.00 |
| Total | 111575 | 88229 | 199804 |
| | 55.84 | 44.16 | 100.00 |
| Priors | 0.5 | 0.5 | |

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.4028 | 0.5052 | 0.4540 |
| Priors | 0.5000 | 0.5000 | |

## Appendix I: SAS Output for Using Quadratic Discriminant Function on Age Group

Using Quadratic Discriminant Function

The DISCRIM Procedure

| | | | |
|---|---|---|---|
| Total Sample Size | 199804 | DF Total | 199803 |
| Variables | 2 | DF Within Classes | 199802 |
| Classes | 2 | DF Between Classes | 1 |

| | |
|---|---|
| Number of Observations Read | 199804 |
| Number of Observations Used | 199804 |

Class Level Information

| age1 | Variable Name | Frequency | Weight | Proportion | Prior Probability |
|---|---|---|---|---|---|
| 0 | _0 | 115580 | 115580 | 0.578467 | 0.500000 |
| 1 | _1 | 84224 | 84224 | 0.421533 | 0.500000 |

Within Covariance Matrix Information

| age1 | Covariance Matrix Rank | Natural Log of the Determinant of the Covariance Matrix |
|---|---|---|
| 0 | 2 | 4.31550 |
| 1 | 2 | 4.82505 |

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV_j^{-1} \, (\bar{X}_i - \bar{X}_j) + \ln |COV_j|$$

Generalized Squared Distance to age1

| From age1 | 0 | 1 |
|---|---|---|
| 0 | 4.31550 | 4.87527 |
| 1 | 4.38737 | 4.82505 |

The SAS System 09:34 Thursday, June 14, 2012 11878

Using Quadratic Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j|$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 93042 | 22538 | 115580 |
| | 80.50 | 19.50 | 100.00 |
| 1 | 58908 | 25316 | 84224 |
| | 69.94 | 30.06 | 100.00 |
| Total | 151950 | 47854 | 199804 |
| | 76.05 | 23.95 | 100.00 |
| Priors | 0.5 | 0.5 | |

Classification Summary for Calibration Data: WORK.ANALY1
Resubstitution Summary using Quadratic Discriminant Function

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.1950 | 0.6994 | 0.4472 |
| Priors | 0.5000 | 0.5000 | |

The SAS System 09:34 Thursday, June 14, 2012 11880

Using Quadratic Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.ANALY1
Cross-validation Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)j}^{-1} \, (X-\bar{X}_{(X)j}) + \ln |COV_{(X)j}|$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 93039 | 22541 | 115580 |
| | 80.50 | 19.50 | 100.00 |
| 1 | 58913 | 25311 | 84224 |
| | 69.95 | 30.05 | 100.00 |
| Total | 151952 | 47852 | 199804 |
| | 76.05 | 23.95 | 100.00 |
| Priors | 0.5 | 0.5 | |

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.1950 | 0.6995 | 0.4473 |
| Priors | 0.5000 | 0.5000 | |

## APPENDIX J: Sas Output Linear Discriminant Analysis of Genotype with Morphometric of each eye

```
The SAS System                        13:53 Friday, June 15, 2012  24
                                          Using Linear Discriminant Function

                                                The DISCRIM Procedure

                          Total Sample Size      110          DF Total               109
                          Variables               10          DF Within Classes      108
                          Classes                  2          DF Between Classes       1


                                    Number of Observations Read            110
                                    Number of Observations Used            110


                                                Class Level Information

                        Variable                                                    Prior
                genotype1    Name        Frequency       Weight    Proportion    Probability

                        0    _0                 50      50.0000      0.454545       0.500000
                        1    _1                 60      60.0000      0.545455       0.500000


                                      Pooled Covariance Matrix Information

                                                     Natural Log of the
                                       Covariance    Determinant of the
                                      Matrix Rank     Covariance Matrix

                                               10             -32.29224


                                 Pairwise Generalized Squared Distances Between Groups

                                           2                 _   _          -1   _   _
                                          D (i|j) = (X  - X )' COV   (X  - X )
                                                      i   j           i   j


                                    Generalized Squared Distance to genotype1

                                 From genotype1             0             1

                                              0             0      20.18256
                                              1      20.18256             0
```

The SAS System 13:53 Friday, June 15, 2012 26

Using Linear Discriminant Function

The DISCRIM Procedure

Linear Discriminant Function

$$\text{Constant} = -.5\ \bar{X}_j' \, COV^{-1} \, \bar{X}_j \qquad \text{Coefficient Vector} = COV^{-1} \, \bar{X}_j$$

Linear Discriminant Function for genotype1

| Variable | Label | 0 | 1 |
|---|---|---|---|
| Constant | | -6101 | -6141 |
| AreaShape_Area | AreaShape_Area | 10.81390 | 11.24510 |
| AreaShape_Eccentricity | AreaShape_Eccentricity | 895.38013 | 913.10132 |
| Areafift | Areafift | 86.06049 | 83.33826 |
| Areatwentyfift | Areatwentyfift | -22.16393 | -20.89337 |
| Areasevenfive | Areasevenfive | -2.12302 | -2.36669 |
| Areaninetyfive | Areaninetyfive | -5.47159 | -5.64012 |
| EccentFive | EccentFive | 1820 | 1813 |
| Eccenttwentyfive | Eccenttwentyfive | -44.23318 | 160.35498 |
| Eccentseventyfive | Eccentseventyfive | -5790 | -6039 |
| eccentninetyfive | eccentninetyfive | 12047 | 12308 |

Using Linear Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = exp(-.5\, D_j^2(X)) \,/\, SUM_k \, exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>100.00 | 0<br>0.00 | 50<br>100.00 |
| 1 | 1<br>1.67 | 59<br>98.33 | 60<br>100.00 |
| Total | 51<br>46.36 | 59<br>53.64 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

Using Linear Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Linear Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0000 | 0.0167 | 0.0083 |
| Priors | 0.5000 | 0.5000 | |

Using Linear Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>100.00 | 0<br>0.00 | 50<br>100.00 |
| 1 | 2<br>3.33 | 58<br>96.67 | 60<br>100.00 |
| Total | 52<br>47.27 | 58<br>52.73 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

Using Linear Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Cross-validation Summary using Linear Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0000 | 0.0333 | 0.0167 |
| Priors | 0.5000 | 0.5000 | |

**APPENDIX K: Sas Output for Quadratic Discriminant Analysis of Genotype with Morphometric of each eye**

```
                                        The SAS System                        13:53 Fri-
day, June 15, 2012  31
                          Using Quadratic Discriminant Function Genotype

                                     The DISCRIM Procedure

                     Total Sample Size      110          DF Total               109
                     Variables               10          DF Within Classes      108
                     Classes                  2          DF Between Classes       1


                           Number of Observations Read            110
                           Number of Observations Used            110


                                     Class Level Information

                        Variable                                                  Prior
            genotype1   Name        Frequency       Weight    Proportion    Probability

                    0   _0                 50      50.0000      0.454545       0.500000
                    1   _1                 60      60.0000      0.545455       0.500000


                              Within Covariance Matrix Information

                                                    Natural Log of the
                                     Covariance    Determinant of the
                        genotype1   Matrix Rank     Covariance Matrix

                                0            10             -35.07331
                                1            10             -40.12572

                          Using Quadratic Discriminant Function Genotype

                                     The DISCRIM Procedure

                     Pairwise Generalized Squared Distances Between Groups

                          2                 _   _          -1  _   _
                         D (i|j) = (X  - X )' COV    (X  - X ) + ln |COV |
                                     i   j      j     i   j           j


                           Generalized Squared Distance to genotype1

                        From genotype1             0             1

                                     0     -35.07331      -3.43804
                                     1      70.04937     -40.12572
```

13:53 Friday, June 15, 2012 33

Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j|$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>100.00 | 0<br>0.00 | 50<br>100.00 |
| 1 | 0<br>0.00 | 60<br>100.00 | 60<br>100.00 |
| Total | 50<br>45.45 | 60<br>54.55 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Quadratic Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0000 | 0.0000 | 0.0000 |
| Priors | 0.5000 | 0.5000 | |

Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Cross-validation Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)j}^{-1} \, (X-\bar{X}_{(X)j}) + \ln |COV_{(X)j}|$$

Posterior Probability of Membership in Each genotype1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into genotype1

| From genotype1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>100.00 | 0<br>0.00 | 50<br>100.00 |
| 1 | 0<br>0.00 | 60<br>100.00 | 60<br>100.00 |
| Total | 50<br>45.45 | 60<br>54.55 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

Using Quadratic Discriminant Function Genotype

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Cross-validation Summary using Quadratic Discriminant Function

Error Count Estimates for genotype1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0000 | 0.0000 | 0.0000 |
| Priors | 0.5000 | 0.5000 | |

**APPENDIX L: Sas Output Linear Discriminant Analysis of Age group with Morphometric measure of each eye**

```
                                        The SAS System                        13:53 Fri-
day, June 15, 2012  37
                                 Using Linear Discriminant Function

                                      The DISCRIM Procedure

                     Total Sample Size      110          DF Total               109
                     Variables               10          DF Within Classes      108
                     Classes                  2          DF Between Classes       1


                              Number of Observations Read            110
                              Number of Observations Used            110


                                    Class Level Information

                         Variable                                                Prior
                age1     Name        Frequency       Weight    Proportion    Probability

                   0     _0                 54      54.0000      0.490909       0.500000
                   1     _1                 56      56.0000      0.509091       0.500000


                                Pooled Covariance Matrix Information

                                                Natural Log of the
                                 Covariance    Determinant of the
                                Matrix Rank     Covariance Matrix

                                          10             -31.61814

                                 Using Linear Discriminant Function


                       Pairwise Generalized Squared Distances Between Groups

                                 2                 _   _          -1   _   _
                                D (i|j) = (X  - X )' COV   (X  - X )
                                                   i   j           i   j


                                Generalized Squared Distance to age1

                              From age1             0             1

                                     0             0       8.27722
                                     1       8.27722             0
```

The SAS System 13:53 Friday, June 15, 2012 39

Using Linear Discriminant Function

Linear Discriminant Function

$$\text{Constant} = -.5\ \bar{X}_j'\ COV^{-1}\ \bar{X}_j \qquad \text{Coefficient Vector} = COV^{-1}\ \bar{X}_j$$

Linear Discriminant Function for age1

| Variable | Label | 0 | 1 |
|---|---|---|---|
| Constant | | -6266 | -6338 |
| AreaShape_Area | AreaShape_Area | 11.10810 | 11.25201 |
| AreaShape_Eccentricity | AreaShape_Eccentricity | 694.53747 | 661.24961 |
| Areafift | Areafift | 92.67307 | 93.32392 |
| Areatwentyfift | Areatwentyfift | -28.61906 | -29.54304 |
| Areasevenfive | Areasevenfive | -0.08297 | 0.24190 |
| Areaninetyfive | Areaninetyfive | -5.36973 | -5.38605 |
| EccentFive | EccentFive | 1781 | 1773 |
| Eccenttwentyfive | Eccenttwentyfive | -491.27369 | -530.99701 |
| Eccentseventyfive | Eccentseventyfive | -4744 | -4603 |
| eccentninetyfive | eccentninetyfive | 11632 | 11611 |

Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)'\ COV^{-1}\ (X-\bar{X}_j)$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\ D_j^2(X))\ /\ SUM_k \exp(-.5\ D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>92.59 | 4<br>7.41 | 54<br>100.00 |
| 1 | 3<br>5.36 | 53<br>94.64 | 56<br>100.00 |
| Total | 53<br>48.18 | 57<br>51.82 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Linear Discriminant Function

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0741 | 0.0536 | 0.0638 |
| Priors | 0.5000 | 0.5000 | |

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 49 | 5 | 54 |
| | 90.74 | 9.26 | 100.00 |
| 1 | 6 | 50 | 56 |
| | 10.71 | 89.29 | 100.00 |
| Total | 55 | 55 | 110 |
| | 50.00 | 50.00 | 100.00 |
| Priors | 0.5 | 0.5 | |

Cross-validation Summary using Linear Discriminant Function

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0926 | 0.1071 | 0.0999 |
| Priors | 0.5000 | 0.5000 | |

**APPENDIX M: Sas Output Quadratic Discriminant Analysis of Age group with Morphometric measure of each eye**

The SAS System
Using Quadratic Discriminant Function

The DISCRIM Procedure

| | | | |
|---|---|---|---|
| Total Sample Size | 110 | DF Total | 109 |
| Variables | 10 | DF Within Classes | 108 |
| Classes | 2 | DF Between Classes | 1 |

| | |
|---|---|
| Number of Observations Read | 110 |
| Number of Observations Used | 110 |

Class Level Information

| age1 | Variable Name | Frequency | Weight | Proportion | Prior Probability |
|---|---|---|---|---|---|
| 0 | _0 | 54 | 54.0000 | 0.490909 | 0.500000 |
| 1 | _1 | 56 | 56.0000 | 0.509091 | 0.500000 |

Within Covariance Matrix Information

| age1 | Covariance Matrix Rank | Natural Log of the Determinant of the Covariance Matrix |
|---|---|---|
| 0 | 10 | -43.58733 |
| 1 | 10 | -30.32510 |

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV_j^{-1} \, (\bar{X}_i - \bar{X}_j) + \ln |COV_j|$$

Generalized Squared Distance to age1

| From age1 | 0 | 1 |
|---|---|---|
| 0 | -43.58733 | -21.61576 |
| 1 | -20.88061 | -30.32510 |

Using Quadratic Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j|$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 52<br>96.30 | 2<br>3.70 | 54<br>100.00 |
| 1 | 4<br>7.14 | 52<br>92.86 | 56<br>100.00 |
| Total | 56<br>50.91 | 54<br>49.09 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Resubstitution Summary using Quadratic Discriminant Function

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0370 | 0.0714 | 0.0542 |
| Priors | 0.5000 | 0.5000 | |

Using Quadratic Discriminant Function

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Cross-validation Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)j}^{-1} \, (X-\bar{X}_{(X)j}) + \ln |COV_{(X)j}|$$

Posterior Probability of Membership in Each age1

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 51<br>94.44 | 3<br>5.56 | 54<br>100.00 |
| 1 | 5<br>8.93 | 51<br>91.07 | 56<br>100.00 |
| Total | 56<br>50.91 | 54<br>49.09 | 110<br>100.00 |
| Priors | 0.5 | 0.5 | |

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WIN1
Cross-validation Summary using Quadratic Discriminant Function

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.0556 | 0.0893 | 0.0724 |
| Priors | 0.5000 | 0.5000 | |

# Appendix N : SAS Output for Validated and training data of cells

## Linear discriminant analysis for Genotype

```
                    Observation Profile for Test Data

            Number of Observations Read          99902
            Number of Observations Used          99902


       Number of Observations and Percent Classified into genotype1

        From genotype1            0            1        Total

                     0        53960        45942        99902
                              54.01        45.99       100.00

                 Total        53960        45942        99902
                              54.01        45.99       100.00

            Priors          0.5          0.5


        Classification Summary using Linear Discriminant Function

                   Error Count Estimates for genotype1

                                         0       Total

                   Rate              0.4599      0.4599
                   Priors            0.5000      0.5000
```

## Quadratic discriminant analysis for Genotype

```
Observation Profile for Test Data

            Number of Observations Read          99902
            Number of Observations Used          99902


       Number of Observations and Percent Classified into genotype1

        From genotype1            0            1        Total

                     0        50039        49863        99902
                              50.09        49.91       100.00

                 Total        50039        49863        99902
                              50.09        49.91       100.00

            Priors          0.5          0.5


           Error Count Estimates for genotype1

                                         0       Total

           Rate              0.4991      0.4991
           Priors            0.5000      0.5000
```

## Linear discriminant analysis for /AGE

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 32931 | 20064 | 52995 |
| | 62.14 | 37.86 | 100.00 |
| 1 | 25775 | 21132 | 46907 |
| | 54.95 | 45.05 | 100.00 |
| Total | 58706 | 41196 | 99902 |
| | 58.76 | 41.24 | 100.00 |
| Priors | 0.5 | 0.5 | |

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.3786 | 0.5495 | 0.4640 |
| Priors | 0.5000 | 0.5000 | |

## Quadratic discriminant analysis for Age

Number of Observations and Percent Classified into age1

| From age1 | 0 | 1 | Total |
|---|---|---|---|
| 0 | 43131 | 9864 | 52995 |
| | 81.39 | 18.61 | 100.00 |
| 1 | 34011 | 12896 | 46907 |
| | 72.51 | 27.49 | 100.00 |
| Total | 77142 | 22760 | 99902 |
| | 77.22 | 22.78 | 100.00 |
| Priors | 0.5 | 0.5 | |

Error Count Estimates for age1

| | 0 | 1 | Total |
|---|---|---|---|
| Rate | 0.1861 | 0.7251 | 0.4556 |
| Priors | 0.5000 | 0.5000 | |