Ain Shams University
Faculty of Commerce
Department of Statistics, Mathematics and Insurance

A Proposed Statistical Model to Study the factors affecting the Recurrence of an Individual's stroke by Myocardial Infarction with Application on Ain Shams University Hospitals

A thesis submitted in partial fulfillment of the requirement for the Master's Degree in Applied Statistics

Presented by

Abeer Ragheb Azazy Hassan
Faculty of Commerce, Ain Shams University

Under the Supervision of

Prof. Dr. Amr Ibrahim Abdelrahman Elatraby
Professor- Department of Statistics, Mathematics and Insurance
Faculty of Commerce, Ain Shams University

Dr. Dina Hassan Abd elhady
Assistant Professor of Applied Statistics,
Faculty of Commerce, Tanta

2013

Approval Sheet

Title of Thesis : A Proposed Statistical Model to Study the factors affecting the Recurrence of an Individual's stroke by Myocardial Infarction with Application on Ain Shams University Hospitals.

Academic Degree : MBA in Applied Statistics.

Name of Student : Abeer Ragheb Azazy Hassan.

This thesis submitted in partial fulfillment of the requirement for

The Master's Degree in Applied Statistics has been approved by:

Examination Committee

1- Prof.Dr. Sohair Fahmy Hegazy

Professor- Department of Statistics
Faculty of Commerce
Tanta University

……

2- Prof.Dr. Amr Ibrahim Abdelrahman Elatraby

Professor- Department of Statistics, Mathematics and Insurance
Faculty of Commerce
Ain Shams University

……

3- Dr. Tolba Elsayed Zein Eldin

Associate Professor- Department of Statistics, Mathematics and Insurance
Faculty of Commerce
Ain Shams University

……

Date of Dissertation Defense / / 2013 Approval Date / / 2013

Acknowledgments

After the help of Allah, many people have been a part of my gratitude; I feel deep gratitude; I feel deep gratitude to those people who have helped me: family, professors, and colleagues. ***Prof. Amr Elatraby****, first, and foremost, has been the best advisor, I could have wished for, he was a great help for all his students. I am grateful to him for pushing me forward, Time after time. In the personal side, he did not hesitate to invite his students to become an extended part of his family. My gratitude also goes to* ***Dr. Dina Hassan Abd elhady*** *for her great supervision, help and advice.*

Very grateful to **Prof. Sohair Hegazy** *for her valuable comments and for her acceptance to share in the examination committee.*

Many thanks to **Associate Prof. Tolba Zein Eldin;** *for his valuable comments and acceptance to share in the examination committee.*

Avery Special thanks are for ***Dr. Osama Mohie Eldin*** *for his positive encouragement to achieve this study. Special thanks and appreciate for the* ***Dr. Mohammed Osama Anwar****; Al Alson High Institute for providing me the help for this* research*.*

I would like to thank those closest to me, whose presence helped to complete my work. Most of all my family; and especially **my father Ragheb Azazy**, *for his absolute confidence in me, and being always there; helping me to overcome all difficulties***.** *Finally, I was lucky enough to have the help of my family. Life would not be the same without the help of my* **Husband Hesham Mostafa**.

Contents

List Of Tables …….. I
List Of Figures.……. III

Summary …….. IV

Chapter One: Introduction

1.1 Introduction ……... 1
1.2 Review of Literature …….. 7
1.3 Importance of the Study …… 12
1.4 Objectives of the study ……..... 13
1.5 Thesis Outlines ……..……. 14

Chapter Two: Categorical Data Analysis

2.1 Introduction ……... 16
2.2 Defining Categorical Variables ……. 19
2.3 Types of Categorical Variables ……. 20
2.4 Categorical Variables properties ……...…… 22
2.5 Methods for Analyzing …….. 23

Chapter Three: Binary Logistic Regression

3.1 Introduction ……..……. 28
3.2 Estimating and Interpretation of the Model ……..…….. 35
3.2.1 Maximum Likelihood Estimation ….. 35
3.2.2 Interpretation the Model Coefficients ……. 38
3.3 Assessing the Fit of the Model ……..... 41
3.3.1 Classification Tables ……..…….. 41
3.3.2 Hosmer-Lemeshow Statistic ……..... 43
3.3.3 Cross Validation Techniques …….. 44
3.3.4 ROC Curve ……. 47
3.3.5 Pseudo R-Squared ……... 50
3.4 Testing for the Significance of the Coefficients ……. 52

3.4.1 Wald Test …… 52
3.4.2 The Likelihood Ratio Test …….. 53
3.5 Confidence Interval Estimation ……..…….. 55

Chapter Four: Linear Discriminant Analysis

4.1 Introduction ……... 57
4.2 Estimating and Interpretation Model ……. 59
4.3 Assessing the Fit of the Model …….. 63
4.3.1 The Eiegenvalues …… 63
4.3.2 Cross Validation and ROC Curve …….. 65
4.4 Testing for the Significance of the Coefficients 66
4.4.1 Wilk's Lambda Test ……..……... 66

Chapter Five: Applied Study

5.1 Data of Binary Logistic Regression Analysis …… 68
5.1.1 Full Model Analysis ……... 71
5.1.2 Stepwise Model Analysis ……... 83
5.2 Data of Linear Discriminant Analysis ……... 90
5.2.1 Full Model Analysis ……..…….. 90
5.2.2 Stepwise Model Analysis ……... 97
5.3 A Comparison between Binary Logistic Model Regression Results and Linear Discriminant Analysis Results …….. 101
5.4 Summary and Conclusions …….…… 102
5.5 Recommendations For Future Research …… 103

Appendices

Appendix A: Binary Logistic Regression Analysis SPSS Output ... 104
Appendix B: Linear Discriminant Analysis SPSS Output …….. 117
Appendix C: Second Myocardial Infarction Survey Questionnaire. 135

References ……... 139

Arabic Summary

List of Tables

Table	Subject	Page
Table (1)	Code sheet for the variables in the study ……...	69
Table (2)	The estimated coefficients and its S.E …….....	71
Table (3)	Odds Ratios and 95% Confidence Intervals for Covariates……..	73
Table (4)	Classification table ……..	75
Table (5)	Classification matrix ……	76
Table (6)	Pseudo R-Squared ……	78
Table (7)	Wald Test …….	79
Table (8)	Likelihood Ratio Test……..	81
Table (9)	Bootstrap for Variables in the Equation …….	82
Table (10)	The estimated coefficients and its S.E and odds ratios…….	83
Table (11)	Classification table …….	85
Table (12)	Classification matrix ……..……..	86
Table (13)	Pseudo R-Squared ……	88
Table (14)	Wald Test …….……	88

Table	Subject	Page
Table (15)	Bootstrap for Variables in the Equation ..……..	89
Table (16)	The estimated Fisher's linear Discriminant functions coefficients……..	91
Table (17)	Classification table ……	92
Table (18)	Classification matrix ……	93
Table (19)	Significance of the Discriminant function……..	95
Table (20)	Bootstrap for Variables in the Equation ……	95
Table (21)	Classification Function Coefficients…….	97
Table (22)	Classification table …….	97
Table (23)	Significance of the Discriminant function …….	100
Table (24)	Bootstrap for Variables in the Equation ……	100
Table (25)	A Comparison between Binary Logistic Regression Results and the Linear Discriminant Analysis Results……	101

List of Figures

Figure	Subject	Page
Figure (1)	Second Myocardial Infarction: Etiology ……	6
Figure (2)	Classification or regression models according to the number of levels of the dependent variable……..	17
Figure (3)	The logistic relationship between dependent and independent variables……...	31
Figure (4)	ROC Curve…….....	49
Figure (5)	Hypothetical frequency distributions of two populations showing percentage of cases incorrectly classified……	60
Figure (6)	Discriminant Analysis with two Groups.....	61
Figure (7)	ROC Curve, Logistic regression full model ……	78
Figure (8)	ROC Curve, Stepwise Binary Logistic regression ….	87