

IDENTIFICATION OF RISK FACTORS FOR TYPE II DIABETES IN ALMADINAH  
ALMUNAWARA, THE KINGDOM OF SAUDI ARABIA

A THESIS  
SUBMITTED TO THE GRADUATE SCHOOL  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS  
FOR THE DEGREE  
MASTRE OF SCIENCE

By

MARAM AHMAD ALRSHEDY

DR. MUNNI BEGUM - ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2016

IDENTIFICATION OF RISK FACTORS FOR TYPE II DIABETES IN ALMADINAH

ALMUNAWARA, THE KINGDOM OF SAUDI ARABIA

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

By

MARAM AHMAD ALRSHEDY

Committee Approval:

.....  
Committee Chairman

Date

.....  
Committee Member

Date

.....  
Committee Member

Date

Department Approval:

.....  
Department Chairperson

Date

.....  
Dean of Graduate School

Date

Ball State University  
Muncie, Indiana  
May 2016

## ACKNOWLEDGMENT

I want to express my deepest gratitude to my advisor Dr. Munni Begum for supporting and encouraging me during my study at Ball State University, and during the period of writing my thesis. She is one of best professors I have ever met, and without her it would not be possible for me to finish my thesis.

I would like to thank my other thesis committee members, Dr. Rebecca Pierce and Dr. Xaio Yayuan for their advice and helpful comments.

I am also appreciative of my dear husband, Fhaid Alrashidi, for standing by my side. It was very difficult to write my thesis, while also taking care of my daughter and son, but without his help, it would have been impossible. I am very thankful to God for bringing him in my life.

I would also like to thank my family for supporting and encouraging me to be successful and to believe in myself.

Finally, a special thanks to my sweet friend Nuha Alsalem, who is always willing to offer help when I need her. She is such a special friend.

Maram Alrshedy

May 2016

## Table of Contents

<b>CHAPTER 1</b>	1
<i>INTRODUCTION</i>	
1.1 The Definition of Diabetes	1
1.2 Diabetes in the World	1
1.3 Diabetes in Saudi Arabia	2
<b>CHAPTER 2</b>	3
<i>LITERATURE REVIEW</i>	3
2.1 Findings from Previous Studies on Diabetes	3
<b>CHAPTER 3</b>	6
<i>METHDOLOGY</i>	6
3.1 The Data	6
3.2 Exploratory Data Analysis	6
3.3 Graphical Summaries	6
3.4 Models for Multinomial Responses	7
3.5 Multinomial Logit Model	8
<b>CHAPTER 4</b>	10
<i>RESULTS AND DISCUSSION</i>	10
4.1 The Exploratory Analysis	10
4.2 Number of Diabetes Patients According to Independent Features	17
4.3 Analysis of Fitting Multinomial Logit Model	19
<b>CHAPTER 5</b>	24
<i>CONCLUSION</i>	24
<b>REFERENCES</b>	25
<b>APPENDIX</b>	27

## List of Tables

### CHAPTER 3

Table 1: The Number of Patients with Four Types of Diabetes by Gender and Nationality in 2009	7
---	---

### CHAPTER 4

Table 2: Characteristics of Patients by Diabetes Type	10
Table 3: Trend in Number of Diabetes Patients	11
Table 4: Characteristic of Patients by Gender	12
Table 5: Characteristic of Patients by Saudi and Non-Saudi Status	13
Table 6: Characteristic of Patients by Type of Diabetes and Gender	14
Table 7: Characteristic of Patients by Type of Diabetes and Year	16
Table 8: Characteristic of Patients by Type of Diabetes and Citizenship	16
Table 9: Characteristic of Patients by Gender & Citizenship	16
Table 10: Characteristic of Patients by Age	17
Table 11: Analyzing No. of diabetes patients for all types of diabetes	17
Table 12: Analyzing No. of diabetes patients for different citizenship status	18
Table 13: Analyzing the number of diabetes patients for male and female	18
Table 14: Analyzing the number of diabetes patients for different age group	19
Table 15: Analysis 1 with Gender and Saudi Status	19
Table 16: Analysis 2 with Age	20
Table 17: Analysis 3 with Time	22

## **List of Figures**

### **CHAPTER 4**

Figure 1: Number of Case According to Diabetes Type	11
Figure 2: Number of Case According to Year	12
Figure 3: Number of Case According to Sex	13
Figure 4: Characteristic of Patients by Citizenship	14
Figure 5: Number of Case According to Diabetes Type and Gender	15

# CHAPTER 1

## INTRODUCTION

### **1.1 The Definition of Diabetes**

Diabetes mellitus or diabetes is defined as a chronic disorder in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond adequately to the insulin that is produced.<sup>1</sup> There are two main types of diabetes: Type-1 diabetes (T1D), which is characterized by the autoimmune destruction of the insulin-producing cells in the pancreas, and Type 2 diabetes (T2D), which is the most common form and is characterized by a reduced production of insulin and an inability of the body tissues to respond fully to insulin.<sup>2</sup>

As there is currently no cure for diabetes, the condition requires lifelong management. In the case of T1D, this means keeping blood glucose levels within safe levels through multiple daily insulin injections or a continuous infusion of insulin through an insulin pump.<sup>3</sup> For T2D, blood glucose levels are managed through medication, diet, and exercise or a combination of these.<sup>3</sup> People with diabetes are also recommended to treatment for lowering cholesterol and blood pressure levels.<sup>3</sup>

### **1.2 Diabetes in the World**

Diabetes represents one of the most challenging public health problems of the 21<sup>st</sup> century and is reaching epidemic levels globally.<sup>4</sup> The total number of people worldwide with T2D was expected to increase from 171 million in 2000 to 366 million in 2030.<sup>4</sup> Unfortunately, the prevalence worldwide already reached 366 million by 2011 according to the international Diabetes Federation (IDF), and the projections are that prevalence of diabetes on a global scale could well reach 530 million people in 2030.<sup>4</sup> Type II diabetes-related mortalities accounted for

4.6 million deaths in 2011 for people aged 20–79 years, accounting for 8.2% of global all-cause mortality for people in this age group with an estimated rate of one death every seven seconds.<sup>5</sup> The number of deaths increased by 13.3% from what had been estimated for the year 2010.<sup>5</sup> The magnitude of the estimated number of deaths due to diabetes is similar to the combined deaths from several infectious diseases, such as HIV/AIDS, malaria, and tuberculosis that are ranked as top public health priorities.<sup>5</sup>

### **1.3 Diabetes in Saudi Arabia**

There is increased concern about the rising tide of T2D and its associated complications in the Arabic speaking countries (East Mediterranean, Arabic peninsula, and Northern Africa) as these regions have some of the highest rates of diabetes in the world.<sup>6</sup> Diabetes prevalence is projected to double over the next two decades in Middle Eastern countries.<sup>7</sup> In December 2011, another alarm awakened Arab governments when International Diabetes Federation announced the latest diabetes estimates at the fifth conference held in Dubai.<sup>1</sup> Six of the top 10 countries with the highest prevalence of diabetes (in adults aged 20 to 79 years) are in the Middle East: Kuwait (21.1%), Lebanon (20.2%), Qatar (20.2%), Saudi Arabia (20.0), Bahrain (19.9%) and UAE (19.2%).<sup>1</sup> In the Arab region, the number of deaths attributed to diabetes is about 170,000 adult people, representing more than 10% of all deaths in the region.<sup>1</sup> However, beyond mortality, temporary and permanent disabilities are often caused by complications of diabetes such as blindness, amputations, kidney failure and cardio vascular diseases (CVDs).<sup>8</sup>

## CHAPTER 2

### LITERATURE REVIEW

#### **2.1 Findings from Previous Studies on Diabetes**

The diabetes pandemic has been much faster than anticipated. The estimated number of diabetic patients worldwide was 171 million in 2000, which is expected to increase to 366 million by 2030.<sup>4</sup> Furthermore, the percentage of diabetics living in developing countries is projected to increase from 74% in 2000 to 81% in 2030.<sup>4</sup> Insulin resistance is believed to be associated with decreased physical activity and obesity. Family history of diabetes, obesity, and hypertension increase the risk of diabetes. Diabetes is also found to be more common among certain ethnic groups. People with pre-diabetes are at a higher risk of developing diabetes, but its onset can be delayed or prevented by dieting, reducing body weight, and increasing physical activity. Unhealthy dietary patterns and lack of exercise are, therefore, the most important factors responsible for the increasing incidence of diabetes worldwide.<sup>9,10</sup>

The epidemiologic transition in the Kingdom of Saudi Arabia (KSA) has been fast and complete. Rapid economic growth during the last 4 decades led to a remarkable increase in living standards and adoption of a ‘Westernized’ lifestyle, characterized by unhealthy dietary patterns, and decreased physical activity.<sup>11</sup> A national survey in 2004 estimated that 23.7% of Saudi adults (age 30-70 years) suffered from T2D, and another 14.1% had impaired fasting glucose.<sup>11</sup> Prevalence of diabetes was significantly higher in urban areas (25.5% versus 19.5% in the rural areas).<sup>12</sup> The burden of diabetes in KSA is likely to increase and reach disastrous levels, unless a comprehensive epidemic control program is implemented rigorously promoting healthy eating, exercise and active lifestyles, and curbing obesity.<sup>13,14</sup> Family history has a major role in the cause of diabetes. Recent studies in genetic research have also identified the genetic variants

linked with T2D.<sup>15,16</sup> Family history of diabetes is also used as a predictor of T2D in population-based screening programs.<sup>17</sup> However, roughly half of the risk of T2D can be attributed to lifestyle, and half to genetics. Lifestyle modification is particularly effective in the prevention, or delay of progression to diabetes among individuals with a family history of diabetes. However, the International Diabetic Federation recommends that diabetes control programs should simultaneously promote lifestyle modification among high-risk individuals, as well as the entire population.<sup>18</sup>

In a 12-year prospective study in the USA, the risk of diabetes significantly increased among men with a ‘Western’ dietary pattern (characterized by higher consumption of red/processed meat, French fries, high-fat dairy products, refined grains, sweets, and desserts), compared to those having a ‘prudent’ dietary pattern comprising of fresh vegetables and fruits, fish, poultry, and whole grains.<sup>19</sup> The risk was significantly greater among obese men.<sup>19</sup> In the USA, poor dietary habits and obesity are closely linked with T2D and its complications.<sup>20</sup> While the Arab populations are known to have a genetic predisposition to diabetes, dietary patterns and physical activity play an equally important role in its cause.

A regional study in Qatar found that obesity, family history, and smoking habits were equally associated with diabetes.<sup>21</sup> In KSA, diabetes, along with hypertension and coronary artery disease has emerged as a major challenge to the health system. The World Health Organization estimates that non-communicable diseases will soon become the principal global cause of morbidity and mortality in KSA.<sup>22</sup>

While the risk factors for T2D are well established, there is very little information on the relationships between different types of T2D and potential risk factors. In Saudi Arabia, no population-based study has been attempted to investigate the association between diabetes types

independently of the effects of gender, age, citizenship status for grouped data. This study attempts to investigate the association between different types of diabetes and demographic risk factors such as, gender, age and citizenship status among patients who live in Saudi Arabia.

## CHAPTER 3

### METHDOLOGY

#### **3.1 Data Exploration**

Data were collected from the primary health care centers (PHCC) of AL Madinah, KSA, from 2009 to 2014. The patients have different types of diabetes: "diabetes during pregnancy & delivery", "diabetes with other complications", "diabetes without complications", or "acetone diabetes." Only a number of demographic factors are available as potential risk factors. These are age groups in 1-4, 5-14, 15-44, and 45+ years, nationality, classified as Saudi and non-Saudi, and gender. Another limitation of the data is that age information is not available for each gender and nationality level. Total number of patients with different types of diabetes is reported for four age groups. Thus we conducted two separate analyses: 1) types of diabetes versus sex and nationality and 2) types of diabetes versus age. Type of diabetes, the response variable is measured on a nominal scale, since there is no natural ordering for the type of diabetes. In our first analysis, gender and nationality are considered as predictors variables and in our second analysis only age is considered as a predictor variable.

#### **3.2 Exploratory Analysis**

Numerical and graphical summaries are created to explore the data and to identify preliminary association between the response and each predictor variable. Exploratory analysis helps to identify particular patterns among the variables in a dataset.

#### **3.3 Graphical Summaries**

The data consist of categorical variables and continuous variables, which are known as qualitative and quantitative variables respectively. We use bar diagrams for categorical variables

and histograms for continuous variables to show the pattern of the variables for the response and the predictors.

### 3.4 Models for Multinomial Responses

When the response categories fall under more than two categories, it can be modeled using the multinomial distribution. The response variable types of diabetes has four categories ; the table below shows the number of patients who have different type of diabetes according to their gender and nationality in 2009.

Table 1: The Number of Patients with Four Types of Diabetes by Gender and Nationality in 2009

(i)		Type (j)				
Covariate Class		1) Diabetes with other Complications	2) Acetone Diabetes or Sugar Com	3) Diabetes during Pregnancy & Delivery	4) Diabetes without Complications	Total
<b>M</b>	<b>Saudi</b>	15453	218	0	6524	22195
<b>M</b>	<b>No</b>	692	19	0	411	1122
<b>F</b>	<b>Saudi</b>	17412	110	2833	5899	26254
<b>F</b>	<b>No</b>	508	13	167	195	883
<b>Total</b>		34065	360	3000	13029	n=50454

The response, which is the type of diabetes, has four categories: “diabetes with other complications”, “acetone diabetes or sugar coma”, “diabetes during pregnancy & delivery” and “diabetes without complications”. Let  $Y_{ij}$  represent one of the possible outcomes, which occurs with probability  $\pi_{ij}$ , representing the count in  $(i, j)^{th}$  cell. For example, in the above table  $Y_{34} = 5899$  which represents the number of cases who have diabetes without complications, and who are Saudi female, has the probability  $\pi_{34}$  with the  $3^{rd}$  covariate class in the  $3^{rd}$  response

category. Then, we have all the possible outcomes that the vector  $Y^T = (y_{11}, y_{12}, \dots, y_{44})$  with probabilities  $(\pi_{11}, \pi_{12}, \dots, \pi_{44})$ , respectively.

In our data the total number of the patients is fixed,  $n = 50454$ , so we cannot consider the Poisson distribution here. However, we can assume that each  $Y_{ij}$  is counted and independent with a Poisson distribution with  $\pi_{ij}$ . Now, we will consider the conditional distribution of all  $Y_{ij}$  on their sum. By using the additive propriety of Poisson distribution, the sum of  $Y_{ij}$  also has a Poisson distribution with parameter  $\pi$ . Then, in our data the joint distribution of all  $Y_{ij}$  on their sum is a multinomial distribution, and it can be written as:

$$f(Y_{i1} = y_{i2}, Y_{i2} = y_{i3}, Y_{i4} = y_{i4} \mid \Sigma Y_{ij} = n) = n! \prod_i \prod_j \frac{\pi_{ij}}{y_{ij}}, \quad i=1, 2, 3, 4, \quad j=1, 2, 3, 4.$$

Thus, the random variable  $Y$  follows the multinomial distribution with mean and variance which can be written as:

$$E(Y) = n\pi, \text{ and } V(Y) = n\pi(1-\pi)$$

### 3.5 Multinomial Logit Model:

The multinomial logit model applies from the binary logistic regression model when the response is categorical and has more than two levels. It is a regression model used to combine predictor variables with a categorical response. The form of the multinomial logit model is similar to the form of the binary logits, but it has more than one logit. The number of logits is one less than the levels of the categorical response. For example, if there are  $k$  response categories, and the  $k^{th}$  categorical response is the reference level, then the multinomial logit model uses  $k-1$  logits which can be written as:

$$\text{Log} ( P(Y = j|x)/P(Y = k|x) ) = X^T \beta_j, \quad j = 1, 2, \dots, k-1$$

Using exponentiation on the both sides of the above equation, we have the following equation:

$$P(Y = j|x) = P(Y = k|x) * \exp (X^T \beta_j)$$

If we take sum on both left and right hand sides over  $j$  on the above expression we get the following equation:

$$\sum_{j=1}^{k-1} P(Y = j|x) = P(Y = k|x) \sum_{j=1}^{k-1} \exp ( X^T \beta_j)$$

Then, by adding  $P(Y= k|x)$  to both sides, we have the following equation:

$$P(Y= k|x) + \sum_{j=1}^{k-1} P(Y = j|x) = P(Y = k|x) \sum_{j=1}^{k-1} \exp ( X^T \beta_j) + P(Y= k|x)$$

$$1 = P(Y= k|x) [1 + \sum_{j=1}^{k-1} \exp ( X^T \beta_j)]$$

By the law of total probability the left hand side of the above equation becomes 1. Thus we have

$$P(Y= k|x) = \frac{1}{1 + \sum_{j=1}^{k-1} \exp ( X^T \beta_j)}$$

And the probability of being in the  $j$  the category becomes:

$$P( Y = j|x) = \frac{\exp (X^T \beta_j)}{1 + \sum_{j=1}^{k-1} \exp (X^T \beta_j)}$$

In our data, the response is the type of diabetes, which has four levels, so the multinomial logit model uses three logits. In addition, we introduce the dummy coding for the independent variables because all of them are categorical variable. So the logits can be written as:

$$\text{Log} ( \pi_1 / \pi_4) = \beta_{01} + \beta_{11} \text{Sex} + \beta_{21} \text{nationality} \quad (1),$$

$$\text{Log} ( \pi_2 / \pi_4) = \beta_{02} + \beta_{12} \text{Sex} + \beta_{22} \text{nationality} \quad (2),$$

$$\text{Log} ( \pi_3 / \pi_4) = \beta_{03} + \beta_{13} \text{Sex} + \beta_{23} \text{nationality} \quad (3).$$

**CHAPTER 4**  
**RESULTS AND DISCUSSION**

**4.1 The Exploratory Analysis**

We have diabetes patients with different kinds of diabetes from 2009 to 2014. The total number of diabetes patients from 2009 to 2014 is 393,241. They are divided into male and female and if they are Saudi or non-Saudi. Also, they have different age ranges.

Table 2: Characteristics of Patients by Diabetes Type

Type	Freq	Cum_Freq	Perc	Cum_Perc
Acetone Diabetes or Sugar Coma	1297	1297	0.33	0.33
Diabetes during Pregnancy & Delivery	11683	12980	2.97	3.3
Diabetes with other Complications	80661	93641	20.51	23.81
Diabetes without Complications	299600	393241	76.19	100

Freq: Frequency counts; Cum\_Freq: Cumulative Frequency; Perc: Percentage, Cum\_Perc: Cumulative percentage

This table shows the number of patients according to the type of diabetes. It shows that the majority of the patients have diabetes without complications. The next largest category is the patients who have diabetes with other complications. The same results can be shown below as a bar plot.

Figure 1: Number of Cases According to Diabetes Type

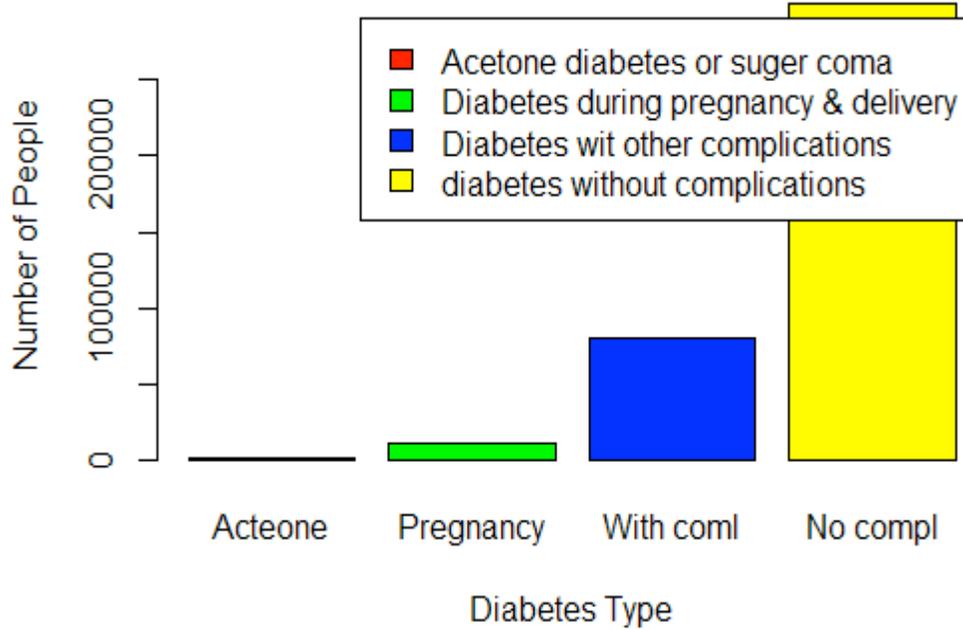


Table 3: Trend in Number of Diabetes Patients

Year	Freq	Cum_Freq	Perc	Cum_Perc
2009	50454	50454	12.83	12.83
2010	25646	76100	6.52	19.35
2011	127780	203880	32.49	51.85
2012	135826	339706	34.54	86.39
2013	47302	387008	12.03	98.41
2014	6233	393241	1.59	100

Freq: Frequency counts; Cum\_Freq: Cumulative Frequency; Perc: Percentage, Cum\_Perc: Cumulative percentage

Figure 2: Number of Cases According to Year

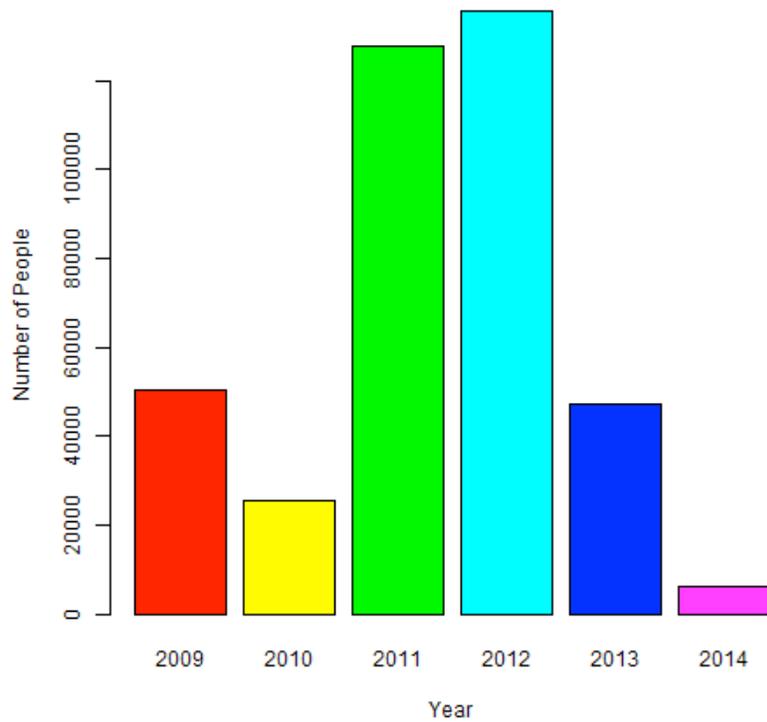


Table 3 and the bar graph in figure 2 show the number of diabetic patients in different years. From 2009 to 2010 the total number of patients decreases, but from 2011 to 2012 it grows gradually. In the last two years the number decreases.

Table 4: Characteristic of Patients by Gender

Gender	Freq	Cum_Freq	Perc	Cum_Perc
Female	223178	223178	56.75	56.75
Male	170063	393241	43.25	100

Freq: Frequency counts; Cum\_Freq: Cumulative Frequency; Perc: Percentage, Cum\_Perc: Cumulative percentage

Table 4 shows that there are 223,178 female and 170,064 male diabetic patients over the years. Women are more likely to have diabetes than men. The same result can be shown in Figure 3.

Figure 3: Number of Cases According to Sex

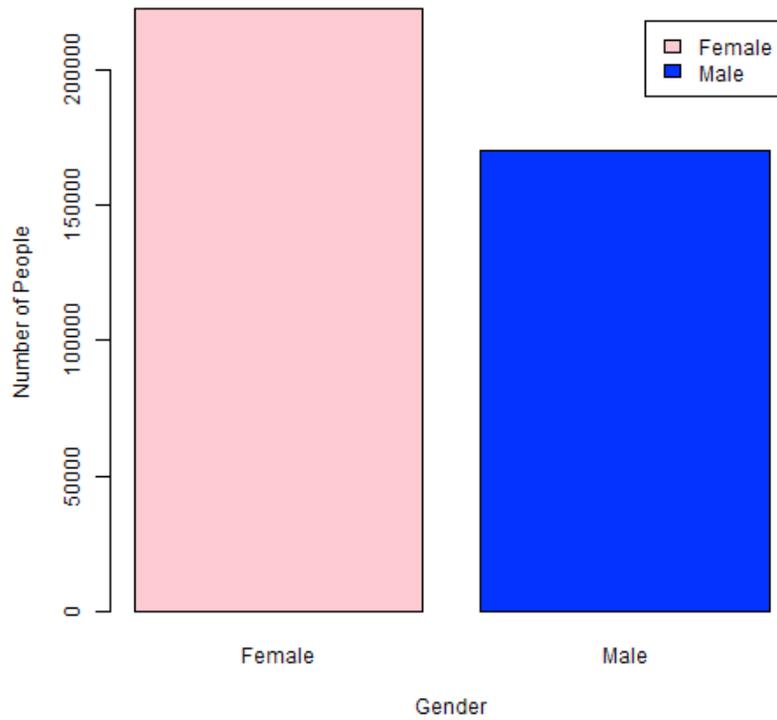
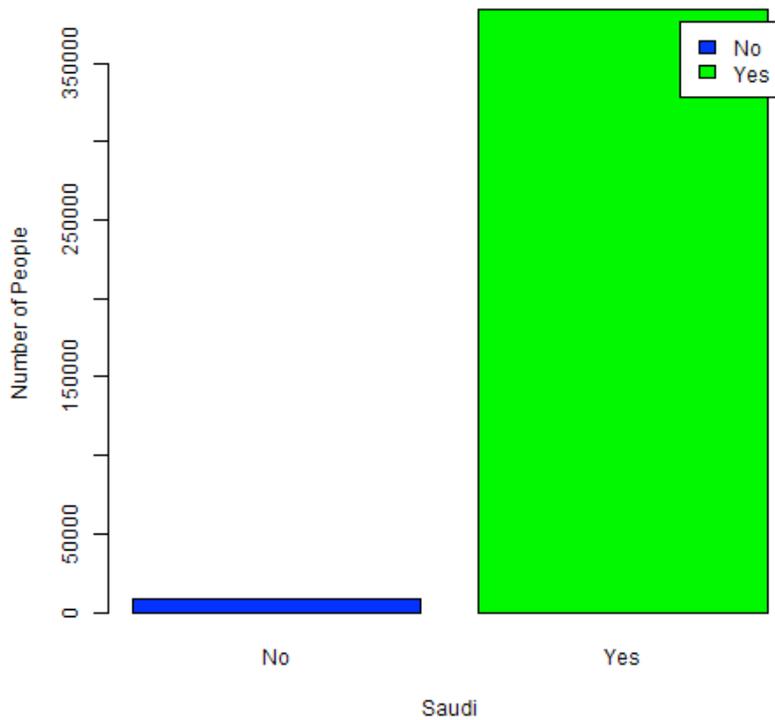


Table 5: Characteristic of Patients by Saudi and Non-Saudi Status

Saudi	Freq	Cum_Freq	Perc	Cum_Perc
Yes	384612	384612	97.80	97.80
No	8629	393241	2.19	100

Freq: Frequency counts; Cum\_Freq: Cumulative Frequency; Perc: Percentage, Cum\_Perc: Cumulative percentage

Figure 4: Characteristic of Patients by Citizenship



Most of the records of diabetes were found in Saudi population. The number is not surprising at all because the survey was in Saudi Arabia, so it is clear that most of them are Saudi.

Table 6: Characteristic of Patients by Type of Diabetes and Gender

	<b>Acetone Diabetes or Sugar Coma</b>	<b>Diabetes during Pregnancy and Delivery</b>	<b>Diabetes with other Complications</b>	<b>Diabetes without Complications</b>
Female	467	11683	45124	165904
Male	830	0	35537	133696

Figure 5: Number of Cases According to Diabetes Type and Gender

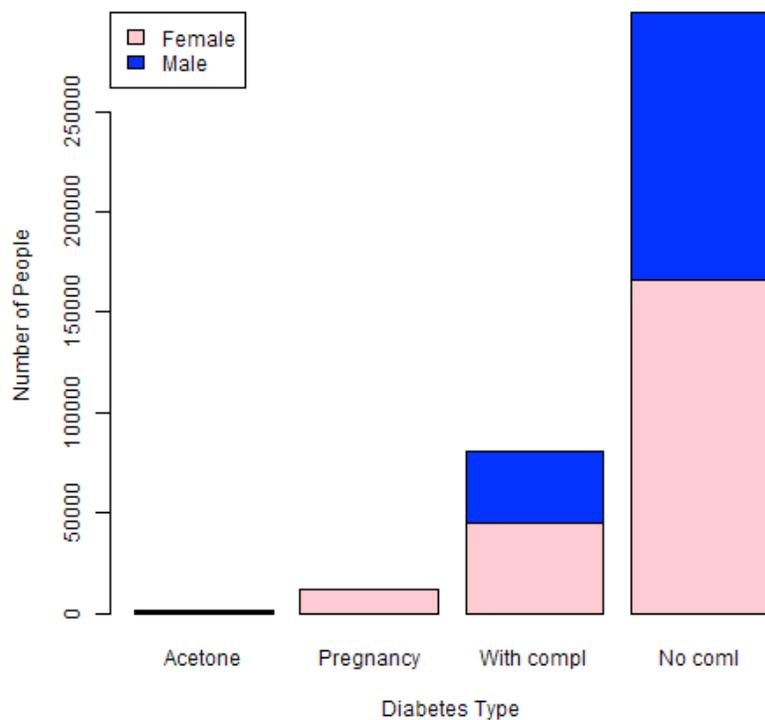


Table 6 and Figure 5 show that most of the patients who have acetone diabetes or sugar coma are men, with 64%. Both men and women have diabetes with other complications with approximately 45% and 56% respectively. This result suggests that both men and women are more likely to have diabetes with other complications.

Table 7: Characteristic of Patients by Type of Diabetes and Year

	2009	2010	2011	2012	2013	2014
Acetone Diabetes or Sugar Coma	360	430	313	155	19	20
Diabetes during Pregnancy & Delivery	3000	2931	1638	1901	1760	453
Diabetes with other Complications	13029	6372	23947	23993	11935	1385
Diabetes without Complications	34065	15913	101882	109777	33588	4375

Table 7 shows the trend of different diabetes in Saudi Arabia. Acetone diabetes and diabetes during pregnancy decrease over the years. However, the total number of patients who have diabetes with complications and without complications increases in 2011 and 2012, then decreases in the last two years.

Table 8: Characteristic of Patients by Type of Diabetes and Citizenship

	Saudi	Non-Saudi
Acetone Diabetes or Sugar Coma	1150	147
Diabetes during Pregnancy & Delivery	11139	544
Diabetes with other Complications	77886	2775
Diabetes without Complications	294437	5163

Table 9: Characteristic of Patients by Gender & Citizenship

	Saudi	Non-Saudi
Female	219552	3626
Male	165060	5003

Table 10: Characteristic of Patients by Age

Age	Freq	Cum_Freq	Perc	Cum_Perc
1 - 4	1296	1296	0.33	0.33
5 - 14	41606	393228	10.58	10.91
15 - 44	137412	138708	34.94	45.85
45+	212914	351622	54.15	100

Freq: Frequency counts; Cum\_Freq: Cumulative Frequency; Perc: Percentage, Cum\_Perc: Cumulative percentage

Most of the diabetic patients are aged 45 or over. About 54% belongs to this group and approximately 35% of the diabetic patients belong to age group 15-44.

#### 4.2 Number of Diabetic Patients According to Independent Features

We are interested in testing several hypotheses.

1. Types of diabetes:

$H_0$ : The number of diabetic patients is same for all types of diabetes.

$H_1$ : The number of diabetic patients is not same for all types of diabetes.

Table 11: Results of analyzing the number of diabetes patients for all types of diabetes

Level of Significance	Test Statistic	Value	Degrees of Freedom	P- Value
0.05	F-Statistic	8.081	3	7.78e-05

Since the p-value is smaller than 0.05, we reject the null hypothesis. Thus, the evidence above indicates that the number of patients is not the same for all types of diabetes.

2. Citizenship status:

$H_0$ : The number of diabetic patients is the same for different citizenship status.

This is equivalent to testing  $H_0: \beta_2 = 0$ , where  $\beta_2$  represents the regression coefficient for nationalism in equations (1), (2), and (3) in section 3.5

$H_1$ : The number of diabetic patients is not same for different citizenship status.

This is equivalent to testing  $H_1: \beta_2 \neq 0$ .

Table 12: Results of analyzing the number of diabetes patients for different citizenship status

Level of Significance	Test Statistic	Value	Degrees of Freedom	P- Value
0.05	F-Statistic	13.75	1	0.000353

Since the p-value is smaller than 0.05, we reject the null hypothesis. Thus, the evidence indicates the number of diabetic patients is not the same for different citizenship status.

### 3. Gender:

$H_0$ : The number of diabetic patients is same for male and female.

This is equivalent to testing  $H_0: \beta_1 = 0$ , where  $\beta_1$  represents the regression coefficient for nationalism in equations (1), (2), and (3) in section 3.5

$H_1$ : The number of diabetic patients is not same for male and female.

This is equivalent to testing  $H_1: \beta_1 \neq 0$ .

Table 13: Results of analyzing the number of diabetes patients for male and female

Level of Significance	Test Statistic	Value	Degrees of Freedom	P- Value
0.05	F-Statistic	0.24	1	0.625

Since the p-value is greater than 0.05, we fail to reject the null hypothesis. Thus, the evidence indicates the number of diabetic patients is the same for males and females.

### 4. Age group:

$H_0$ : The number of diabetic patients is same for the different age groups.

$H_1$ : The number of diabetic patients is not same for the different age groups.

Table 14: Results of analyzing the number of diabetes patients for different age group

Level of Significance	Test Statistic	Value	Degrees of Freedom	P- Value
0.05	F-Statistic	4.031	3	0.00965

Since the p-value is smaller than 0.05, we reject the null hypothesis. Thus, the evidence indicates the number of diabetic patients is not the same for different age groups.

### 4.3 Analysis of Diabetes Types: Multinomial Logit Model

We are interested to determine the effect of gender and nationality on diabetes type. We set up the model by taking the type of diabetes as the dependent variable with gender and Saudi status as the independent variables. The result of fitting the multinomial logit model is given below:

Table 15: Results of analysis with Gender and Saudi Status

	Predictors	Coefficients	Standard Error	Z-value	P-value	Odds ratio
Diabetes during pregnancy & delivery	Intercept	2.784053875	0.10448728	26.64491	0	16.18450
	Saudi Yes	0.4624679	0.09992012	4.628376	0.0000036	1.587988
	GenderMale	-13.0050780	4.68596591	-2.775325	0.00551465	0.0000022
Diabetes with other complications	Intercept	3.470552	0.09591182	36.18482	0	32.15450
	SaudiYes	1.1506672	0.09018960	12.758313	0	3.160301
	GenderMale	-0.7663361	0.05856958	-13.084199	0	0.464712
Diabetes without complications	Intercept	4.069691	0.09495621	42.858612	0	58.53886
	SaudiYes	1.8658343	0.08919134	20.919457	0	6.461324
	GenderMale	-0.7309898	0.05826282	-12.546420	0	0.4814322

## Diabetes Type and Age:

Since the age specific information is not available for gender, nationality, and types of diabetes, but only available for types of diabetes, we perform a separate analysis for diabetes type and age. Results are presented in Table 16.

Table 16: Results of analysis with Age

	<b>Predictors</b>	<b>Coefficients</b>	<b>Standard Error</b>	<b>Z-value</b>	<b>P-value</b>	<b>Odds ratio</b>
Diabetes during pregnancy & delivery	Intercept	-17.124058	0.04560436	-375.491701	0.0	3.6569e-08
	Age 5-14	17.5413630	0.1299650	134.969938	0.0	4.15066e+07
	Age 15-44	19.7265576	0.05501756	358.550236	0.0	3.69092e+8
	Age 45+	18.905974	0.05440135	347.527637	0.0	1.62465e-01
Diabetes with other complications	Intercept	4.646320	0.04560436	14.622911	0.0	1.04200e+02
	Age 5-14	0.4084619	0.3448647	1.184412	0.23624	1.50450
	Age 15-44	-0.8718708	0.32056895	-2.719761	6.533e-03	4.18168e-01
	Age 45+	-0.405584	0.32013745	-1.266906	2.052e-01	6.66587e-01
Diabetes without complications	Intercept	3.194581	0.32264376	9.901264	0.0	2.43999e+01
	Age 5-14	3.1747858	0.3492664	9.089869	0.0	2.39217e+01
	Age 15-44	2.0218746	0.32538016	6.213884	5.169e-10	7.55247e+00
	Age 45+	2.303617	0.32497873	7.088515	1.355e-12	1.00103e+01

Under the multinomial logit model using the R package, we created the table of coefficients and odds ratio (OR) with p-value and z-value. The dependent variable (diabetes type) has four levels, so we fit three multinomial logit models to predict the probability of types of diabetes. In the above table, we set acetone diabetes as the reference group so each model will

give us the probability of being in a particular type of diabetes compared to the reference group. Also, we set a reference group for all our independent variables because they are all categorical variables.

From Table 15, our model equation can be written as:

$$\text{Ln} \left( \frac{p(\text{being in diabetes during pregnancy})}{p(\text{being in acetone diabetes})} \right) = \beta_{10} + \beta_{11}\text{SaudiYes} + \beta_{12}\text{GenderMale} \quad (1)$$

$$\text{Ln} \left( \frac{p(\text{being in diabetes with other complications})}{p(\text{being in acetone diabetes})} \right) = \beta_{20} + \beta_{21}\text{SaudiYes} + \beta_{22}\text{GenderMale} \quad (2)$$

$$\text{Ln} \left( \frac{p(\text{being in diabetes without complications})}{p(\text{being in acetone diabetes})} \right) = \beta_{30} + \beta_{31}\text{SaudiYes} + \beta_{32}\text{GenderMale} \quad (3)$$

We can interpret the result in terms of log odds in equation (1), the log odds of being diabetic during pregnancy compared to acetone diabetes is increasing by 0.46 for Saudi people than non-Saudi. In equation (2), the log odds of being diabetic with other complications compared to acetone diabetes is decreasing by -0.77 for male than female.

In addition, we can interpret the result in terms of odds ratio. For example, the second value of odds ratio column in Table 15, the risk of having diabetes during pregnancy and delivery to acetone diabetes is 1.59 times higher for Saudi people than non-Saudi. Further, the findings suggest that all the variables are statistically significant since the p-values are less than 0.05. For the model with gender and Saudi status all levels of diabetes types are statistically significant compare to acetone diabetes for the predictors.

Similarly, the model equation from Table 16:

$$\text{Ln} \left( \frac{p(\text{being in diabetes during pregnancy})}{p(\text{being in acetone diabetes})} \right) = \beta_{10} + \beta_{11}\text{Age}_{5-14} + \beta_{12}\text{Age}_{15-44} + \beta_{13}\text{Age}_{45+} \quad (4)$$

$$\text{Ln} \left( \frac{p(\text{being in diabetes with complications})}{p(\text{being in acetone diabetes})} \right) = \beta_{20} + \beta_{21}\text{Age}_{5-14} + \beta_{22}\text{Age}_{15-44} + \beta_{23}\text{Age}_{45+} \quad (5)$$

$$\text{Ln} \left( \frac{p(\text{being in diabetes without complications})}{p(\text{being in acetone diabetes})} \right) = \beta_{20} + \beta_{21} \text{Age}_{5-14} + \beta_{22} \text{Age}_{15-44} + \beta_{23} \text{Age}_{45+}$$

(6)

We can interpret the result in terms of log odds in equation (4) the log odds of being in diabetes during pregnancy comparing to acetone diabetes is increasing by 17.54 for patients whose age category is 5 – 14 than patients whose age category is 1 – 4. In equation (5), the log odds of being in diabetes with other complications comparing to acetone diabetes is decreasing by -0.87 for patients whose age category is 15 – 44 than patients whose age category is 1 – 4.

Moreover, we can interpret the result in terms of odds ratio (OR). For example, the 6<sup>th</sup> value of odds ratio column in Table 16, the risk of having diabetes with other complications to acetone diabetes is 1.5045 times higher for people whose age category is 15 – 44 than patients whose age category is 1 – 4. In the multinomial logit model with age variable, the effect of age group 5-14 and 45+ compare to age group 1-4 have same effect for having diabetes with other complications to acetone diabetes. All other levels of diabetes are statistically significant compare to acetone diabetes for all levels of age.

Table 17: Analysis with Time

	Predictors	Coefficients	Standard Error	Z-value	P-value	Odds ratio
Diabetes during pregnancy & delivery	Intercept	101.0294	7.189112e-09	14053113581	0.0	7.525e+43
	Year	-0.04915193	1.445563e-05	-3400.193	0.0	0.9520365
Diabetes with other complications	Intercept	-478.4869	6.874231e-09	-6960588987	0.0	1.569e-21
	Year	0.23997386	1.382463e-05	17358.431	0.0	1.2712159
Diabetes without complications	Intercept	-629.4181	6.835540e-09	-9208024102	0.0	4.4379e-3
	Year	0.31566852	1.374474e-05	22966.494	0.0	1.3711757

In the same way, we can obtain our model equation for table 17:

$$\text{Ln} \left( \frac{p(\text{being in diabetes during pregnancy})}{p(\text{being in acetone diabetes})} \right) = \beta_{10} + \beta_{11} \text{Year} \quad (7)$$

$$\text{Ln} \left( \frac{p(\text{being in diabetes with other complications})}{p(\text{being in acetone diabetes})} \right) = \beta_{20} + \beta_{21} \text{Year} \quad (8)$$

$$\text{Ln} \left( \frac{p(\text{being in diabetes without complications})}{p(\text{being in acetone diabetes})} \right) = \beta_{30} + \beta_{31} \text{Year} \quad (9)$$

With one unit increase in year variable the log odds of being in diabetes during pregnancy comparing to acetone diabetes decrease by 0.049. Also, in the multinomial logit model with year variable all levels of diabetes types are statistically significant compare to acetone diabetes for the predictors as their p-value are less than 0.05.

## CHAPTER 5

### CONCLUSION

From the public health point of view, identification of the risk factors for different types of diabetes is very important. Since the type of diabetes is measured on a nominal scale with four levels, we considered a multinomial logit model to estimate the odds of being in different categories for gender, nationality, and age groups. The Multinomial Logit model allows a nominal categorical response variable to have more than two levels. We used this model to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables. In addition, the model can give us a clear picture about the effect of the independent variables.

The results from the analysis of the type of diabetes and gender and nationality show that the odds of having different types of diabetes differ across gender and nationality. Also the results from the analysis of the type of diabetes and age show that the odds of having different types of diabetes differ across age groups, which is expected.

One limitation of our study is that the data does not include age specific information for gender and nationality. This led us to conduct a separate analysis to determine the association between the type of diabetes and age group.

This study can be extended by considering more covariate or risk factors, if available. A trend component to the model may also improve the predictability of the model. Due to time constraints and limitations of the data, these features were not explored in the current research.

## REFERENCES

1. Boutayeb1 A, Lamlili M, Boutayeb W, Maamri A, Ziyat A, Ramdani N. The rise of diabetes prevalence in the Arab region. *Open J Epidemiol.* 2012; (2): 55-60.  
[http://file.scirp.org/pdf/OJEpi20120200003\\_22690219.pdf](http://file.scirp.org/pdf/OJEpi20120200003_22690219.pdf). Accessed June 20, 2015.
2. Trimarchi C. The Connection between Diabetes and Hearing Loss. Retiree Web site.  
<http://www.the-retiree.com.au/features/the-connection-between-diabetes-and-hearing-loss/>.  
Published December 29, 2012. Accessed June 20, 2015.
3. Shaw J. Diabetes: the silent pandemic and its impact on Australia. Baker IDI Heart and Diabetes Institute. 2012; 1-27.  
[https://www.diabeteswa.com.au/UserDir/Documents/14814%20Diabetes%20management%20booklet%205%20Mar\\_FINAL.pdf](https://www.diabeteswa.com.au/UserDir/Documents/14814%20Diabetes%20management%20booklet%205%20Mar_FINAL.pdf). Accessed July 15, 2015.
4. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004; 27: 1047-1053.
5. International Diabetes Federation, IDF Diabetes Atlas, International Diabetes Federation, Brussels, Belgium, 5th edition, 2011, <http://www.idf.org/diabetesatlas>. Accessed June 22 2015.
6. Alhyas L, McKay A, Balasanthiran A, Majeed A. Prevalences of overweight, obesity, hyperglycaemia, hypertension and dyslipidaemia in the Gulf: systematic review. *J R S M.* 2011; (2): 1-55. <http://shr.sagepub.com/content/2/7/55.short>. Accessed July 13, 2015.
7. Alkhudairy L, Stranges S, Kumar S, Aldaghri N, Rees K. Dietary Factors and Type 2 Diabetes in the Middle East: What Is the Evidence for an Association? A Systematic Review. *PMC.* 2013; 5(10): 3871-3897. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3820049/>. Accessed September 14, 2015.
8. Greene RR. *Contemporary issues of care.* 2<sup>nd</sup> ed. Routledge, NY: The Haworth Press; 2011.
9. World Health organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Report of a WHO/IDF Consultation 2006. Geneva (Switzerland): WHO Publications; 2006. Available from URL:  
[http://www.who.int/entity/diabetes/publications/Definitions/diagnosis/diabetes\\_new.pdf](http://www.who.int/entity/diabetes/publications/Definitions/diagnosis/diabetes_new.pdf)
10. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001; 345: 790-797.

11. Al-Hazzaa HM. Prevalence of physical inactivity in Saudi Arabia: a brief review. *East Mediterr Health J* 2004; 10: 663- 670.
13. Al-Nozha MM, Al-Maatouq MA, Al-Mazrou YY, Al-Harhi SS, Arafah MR, Khalil MZ, et al. Diabetes mellitus in Saudi Arabia. *Saudi Med J* 2004; 25: 1603-1610.
14. Swinburn B. Sustaining dietary changes for preventing obesity and diabetes: lessons learned from the successes of other epidemic control programs. *Asia Pac J Clin Nutr* 2002; 11 (Suppl 3): S598-S606.
15. Hussain A, Claussen B, Ramachandran A, Williams R. Prevention of type 2 diabetes: a review. *Diabetes Res Clin Pract* 2007; 76: 317-326.
16. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 2008; 359: 2220-2232.
17. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; 445: 881-885.
18. Hariri S, Yoon PW, Qureshi N, Valdez R, Scheuner MT, Khoury MJ. Family history of type 2 diabetes: a population-based screening tool for prevention? *Genet Med* 2006; 8: 102-108.
19. Alberti KG, Zimmet P, Shaw J. International Diabetes Federation: a consensus on Type 2 diabetes prevention. *Diabet Med* 2007; 24: 451-463.
20. Van Dam RM, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Dietary patterns and risk for type 2 diabetes mellitus in U.S. men. *Ann Intern Med* 2002; 136: 201-209.
21. Mokdad AH, Serdula MK, Dietz WH, Bowman BA, Marks JS, Koplan JP. The spread of the obesity epidemic in the United States, 1991-1998. *JAMA* 1999; 282: 1519-1522.
22. Bener A, Zirrie M, Janahi IM, Al-Hamaq AO, Musallam M, Wareham NJ. Prevalence of diagnosed and undiagnosed diabetes mellitus and its risk factors in a population-based study of Qatar. *Diabetes Res Clin Pract* 2009; 84: 99-106.
23. Chopra M, Galbraith S, Darnton-Hill I. A global response to a global problem: the epidemic of overnutrition. *Bull World Health Organ* 2002; 80: 952-958.
24. Tutz G. Multinomial Response Models. *Regression for Categorical Data*: 2011: 207–214.

## APPENDIX

### R code:

```
setwd("C:/Users/Maram Alrshedy/Desktop/thesis")
data <- read.table("diabetes.csv", header=TRUE, sep=",")
head(data)
names(data)
str(data)
tail(data)
class(data)      # "data.frame"
sapply(data, class) # show classes of all columns
typeof(data)     # "list"
names(data)      # show list components
dim(data)        # dimensions of object, if any
head(data)       # extract first few (default 6) parts
tail(data, 1)    # extract last row
head(1:10, -1)   # extract everything except the last element
attach(data)
mytable <- table(case,year) # A will be rows, B will be columns
mytable
summary(mytable)
plot(year,case)
mytable
table(case,year)
table(case,gender)
table(case,saudi)
table(case,type)
table(gender)
data[,4=="yes"]="Yes"
#3-way frequency table
mytable<- table(case,gender,year)
ftable(mytable)
summary(mytable)
#3-way frequency table
mytable<- xtabs(~case+gender+year, data=data)
ftable(mytable)
summary(mytable)
#creating a tabal
table_Diabetes=xtabs(case ~ type)
table_Diabetes
table_Year=xtabs(case ~ year)
table_Year
table_Gender=xtabs(case ~ gender)
table_Gender
```

```

table_Citizen=xtabs(case ~ saudi)
table_Citizen
table_Diabetes_Gender=xtabs(case ~ gender+type)
table_Diabetes_Gender
table_Diabetes_Year=xtabs(case ~ type+year)
table_Diabetes_Year
table_Diabetes_Citizen=xtabs(case ~ type+saudi)
table_Diabetes_Citizen
table_Gender_Citizen=xtabs(case ~ gender+saudi)
table_Gender_Citizen
for(i in data$saudi) if(i %in% "yes") data$saudi <- "Yes"
data$Species[data$Species == 'yes'] <- 'Yes'
rename(data, yes = Yes)
colnames(data)[colnames(data)=="No..of.cases"] <- "cases"
library(dplyr)
rename(data, yes = Yes)
rename
rename <- function(data, yes, Yes) {
data <- saudi(data)
data[which(data %in% yes)] <- Yes
saudi(data) <- data
data
table(saudi)
levels(data$saudi)[levels(data$saudi)=="yes"] <- "Yes"
levels(data$saudi)
table(data$saudi)
#creating a barplot
png('barplot1.png')
#creating a barplot
barplot(table_Diabetes,xlab="Diabetes Type",ylab="Number of People",
names.arg=c("with compl", "Acetone", "Pregnancy", "No compl"), main="No. of Case according
to Diabetes Type",
col=c("red", "green", "blue", "yellow"), legend = rownames(table_Diabetes))
dev.off()
png('barplot1.png')
barplot(table_Year,xlab="Year",ylab="Number of People",
main="No. of Case according to Year",col=rainbow(6))
dev.off()
png('barplot1.png')
barplot(table_Gender,xlab="Gender",ylab="Number of People",main="No. of Case according to
Sex",
col=c("pink", "blue"), legend = rownames(table_Gender))
dev.off()
barplot(table_Citizen,xlab="Saudi",ylab="Number of People",main="No. of Case according to
Citizenship",
col=c("blue", "green"), legend = rownames(table_Citizen))

```

```

png('barplot1.png')
barplot(table_Diabetes_Gender, main="No. of Case According to Diabetes Type & Gender ",
xlab="Diabetes Type", ylab="Number of People", col=c("pink","blue"),
names.arg=c("Acetone","Pregnancy","With compl","No compl"));
legend("topleft",c(col=rownames(table_Diabetes_Gender)),fill=c("pink","blue"))
dev.off()
barplot(table_Diabetes_Citizen, main="No. of Case According to Diabetes Type & Citizenship "
xlab="Diabetes Type", ylab="Number of People", col=c("blue","green"),
names.arg=c("Acetone","Pregnancy","With compl","No compl"));
legend("topleft",c(col=rownames(table_Diabetes_Citizen)),fill=c("blue","green"))
# Grouped Bar Plot
counts <- table(mtcars$vs, mtcars$gear)
barplot(table_Diabetes_Gender, main="No. of Case According to Diabetes Type & Gender ",
xlab="Diabetes Type", ylab="Number of People", col=c("darkblue","red"),
names.arg=c("No compl", "Acetone","Pregnancy","Other"),
legend = rownames(table_Diabetes_Gender), beside=TRUE)
library(ggplot2)
ggplot(data=data, aes(x=type, y=case, fill=gender)) +
geom_bar(stat="identity")
ggplot(data=data, aes(x=type, y=case, fill=gender)) +
geom_bar(colour="black", stat="identity",
position=position_dodge(),
size=.3) + # Thinner lines
scale_fill_hue(name="Sex of people") + # Set legend title
xlab("Diabetes Type") + ylab("Number of People") + # Set axis labels
ggtitle("No. of Case According to Diabetes Type & Gender") + # Set title
theme_bw()
#Main Analysis
require(nnet)
logit = multinom(type ~ saudi+gender,weights = case, data=data)
summary(logit)
table(data$saudi)
z <- summary(logit)$coefficients/summary(logit)$standard.errors
#2-tailed z test
p <- (1 - pnorm(abs(z), 0, 1))*2
exp(coef(logit))
logit = multinom(type ~ year,weights = case, data=data)
summary(logit)
exp(coef(logit))
z <- summary(logit)$coefficients/summary(logit)$standard.errors
#2-tailed z test
p <- (1 - pnorm(abs(z), 0, 1))*2
p <- (1 - pnorm(abs(z), 0, 1))*2
exp(coef(logit))
# One Way Anova (Completely Randomized Design)
attach(data)

```

```
m1 <- aov(case ~ type, data=data)
summary(m1)
m2 <- aov(case ~ saudi, data=data)
summary(m2)
m3 <- aov(case ~ gender, data=data)
summary(m3)
m4 <- aov(case ~ year,data=data)
summary(m4)
# Rename a column in R
colnames(data)[colnames(data)=="No..of.cases"] <- "cases"
table_Age=xtabs(cases ~ Age)
table_Age
#Main Analysis
require(nnet)
logit = multinom(Type ~ Age,weights = cases, data=data)
summary(logit)
z <- summary(logit)$coefficients/summary(logit)$standard.errors
#2-tailed z test
p <- (1 - pnorm(abs(z), 0, 1))*2
exp(coef(logit))
attach(data)
m1 <- aov(cases ~ Age, data=data)
summary(m1)
```