ITCS 5102 Survey of Programming Languages - Term Project

# 

# **Team Name : R Challengers**

# 

## Contributors:

## Advaith Auron Suresh(800943085)

## Anirudh Lakshminarayanan(800964326)

## Kandasamy Kishore(800941654)

## Sriram Subramanian(800959758)

## Shiva Vandana Nayakanti(800951756)

# **Contents:**

### **Introduction**

### **R Programming Language**

### **History of R**

### **Major Data Structures**

### **Problem Analysis**

### **Data disparity**

### **Parameter analysis : Attribute selection**

### **Correlation : Race**

### **Correlation : Age**

### **Correlation : Medication use**

### **Correlation : HbA1c result**

### **Analysis Summary**

### **Conclusion**

### **Implementational difficulties**

### **References**

# Introduction:

A healthcare organization estimates that they suffer close to $25 million in preventable losses every year because of readmitting patients who get discharged from the hospital earlier than they should have. Reducing re-hospitalizations among medicare beneficiaries has taken high priority.

Hospital readmissions are seen as an important indicator of care quality and account for billions of dollars in annual Medicare spending. The aim of our project is to develop a model of readmission risk that can be considered by the hospital management when treating a patient.

# 

# Programming Language:

The programming language used to analyse the data proposed in this project is **R programming** language. The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. R allows integration with the procedures written in the C, C++, .Net, Python or FORTRAN languages for efficiency. R programming was used for the following reasons:

* Vast package system
* Provides graphical facilities for data analysis
* Easy to extend functionalities
* Provides operators for calculations on arrays, lists, matrices and vectors
* Effective data handling and storage facility

## 

## History of R programming:

R was initially written by **Ross Ihaka** and **Robert Gentleman** at the Department of Statistics of the University of Auckland in Auckland, New Zealand. R made its first appearance in 1993.

* A large group of individuals has contributed to R by sending code and bug reports.
* Since mid-1997 there has been a core group who can modify the R source code archive.

## Major Data Structures:

In R programming, variables are assigned with R-objects and the data type of the variable depends on the data type of the R-object. The commonly used data structures in R programming are:

* Arrays
* Lists
* Vectors
* Matrices
* Factors

The simplest of these are vectors and the other data structures are built upon these vectors.

Lists are objects that can contain different types of elements such as vectors or functions. Matrices just like the mathematical functionality is a two dimensional data set used to matrix functions also.

Arrays are similar to matrices but can take many dimensions unlike matrices which can take only 2 dimensions. Factors add a distinct value of elements in the vector as labels.

# Problem Analysis:

The initial dataset used contained over 10000 records. We refined the dataset by removing patient entries that resulted in deaths or home discharges by using the following commands:

Patients ← subset(patients, discharge, disposition\_id != “Hospice” )

Patients ←- subset(patients, discharge, disposition\_id != “Expired” )

The dataset reduced by a significant number.

### 

### **Data Disparity:**

Rules were executed in order to determine possible violations on the dataset.

Ex:

library(editrules)

(A← editfile(“rule.txt”)

Violate ← violatedEdits(A, patients)

summary(Violate)

### **Parameter Analysis: Attribute Selection**

The following attributes were used in relation to the readmission rate analysis:

* Race
* Age
* Medication Use
* A1Cresult

#### 

#### **Correlation: Race**

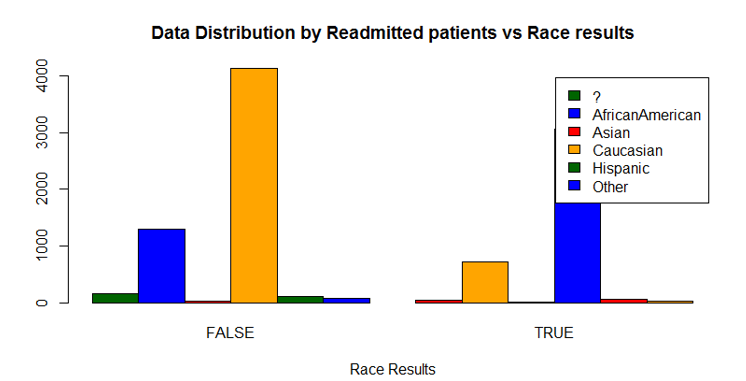
The first criterion considered was Race. The following R commands were run to obtain the correlation with readmission attribute:

>tab <- xtabs(~race + readmitted, data = patients)

>barplot(tab, main="Data Distribution by Readmitted patients vs Race results",xlab="Race Results",col=c("darkgreen","blue","red","orange")

legend = rownames(tab), beside=TRUE)

>summary(assocstats(tab))



Similarly, the following commands were run to obtain the relation between readmitted patients and age, medication change and A1C results.

#### **Correlation: Age**

R commands used:

>tab <- xtabs(~age + readmitted, data = patients)

>barplot(tab, main="Data Distribution by Readmitted patients vs Age results",xlab="Age Results",col=c("darkgreen","blue","red","orange","yellow","maroon","pink","violet","cyan","magenta"), legend = rownames(tab), beside=TRUE)

>summary(assocstats(tab))

#### **Correlation: Different Medication Use**

R commands used:

>tab <- xtabs(~change + readmitted , data = patients)

>barplot(tab, main="Data Distribution by Readmitted patients vs Change in Medication results", xlab="Change in Medication Results", col=c("darkgreen","red"), legend = rownames(tab), beside=TRUE)

>summary(assocstats(tab))

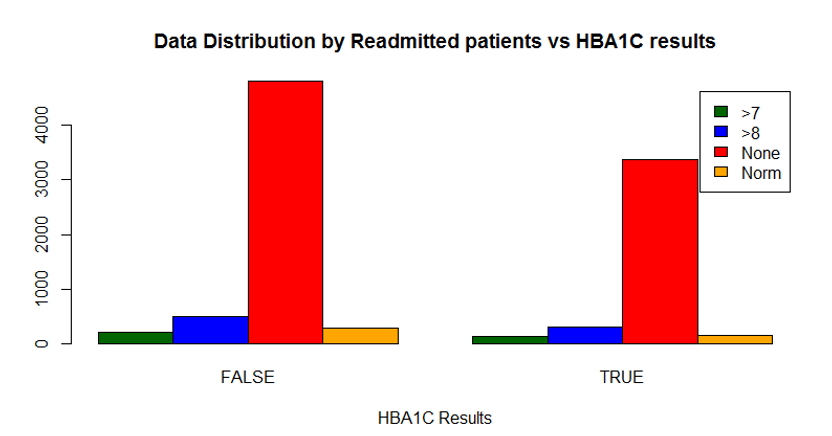
#### **Correlation: HbA1C**

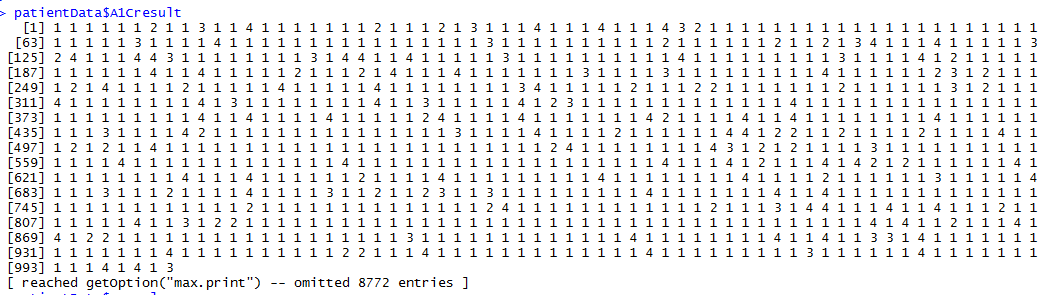
R commands used:

>tab <- xtabs(~A1Cresult + readmitted, data = patients)

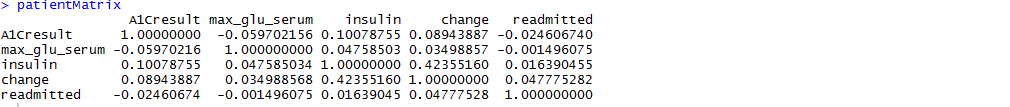
>barplot(tab, main="Data Distribution by Readmitted patients vs HBA1C results",xlab="HBA1C Results", col=c("darkgreen","blue","red","orange"),legend = rownames(tab), beside=TRUE)

>summary(assocstats(tab))

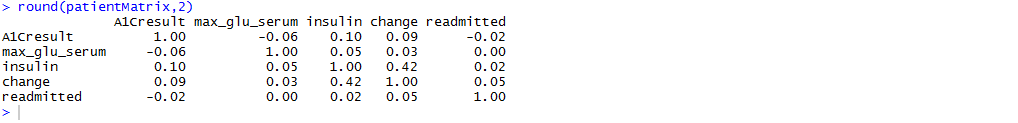




**Correlation of the patient matrix data :**



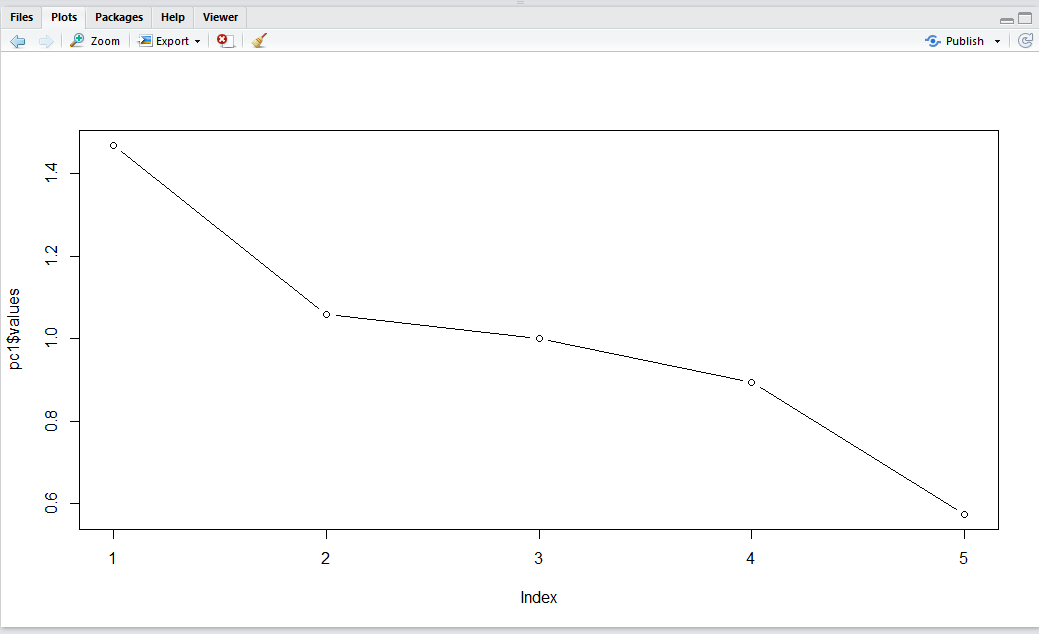
**Rounding Off the patient Matrix data :**



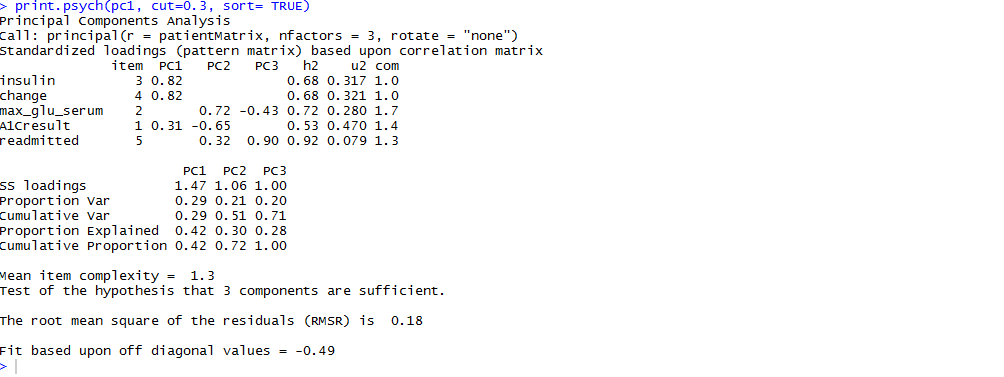
**Calculating Chi Square value :**



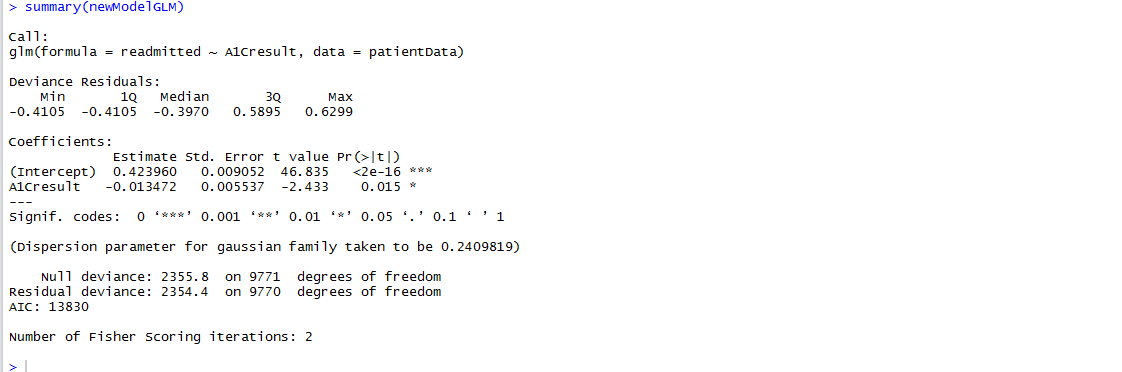
**Plot graph :**



**print(psych):**



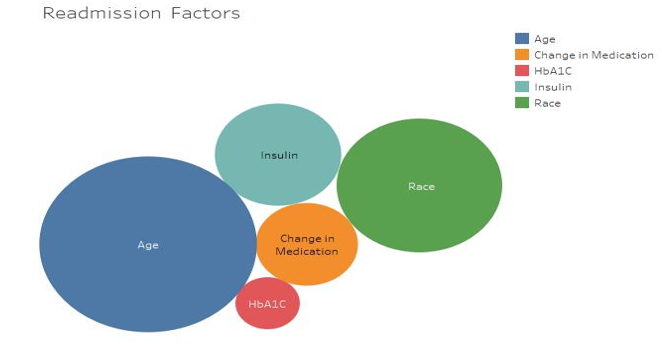
**Summary of Linear Regression Model :**



### 

### **Analysis Summary:**

The analysis of the given parameters allowed us to determine which factor had the most impact on the readmission rate of patients in the hospital. We observed the chi-square values of all the selected parameters and made a comparative study. We found out that the chi-square value for HbA1C test factor that the highest influence on the readmission rate. We have illustrated the same below:



**Fig: Colors indicate factors and chi-squared values indicate impact on readmission rate**

# Conclusion:

We hypothesize that the measurement of HbA1c can be directly correlated with the readmission rates in individuals admitted to the hospital. Analyzing different groups of individuals with different levels of HbA1c content with different scenarios such as the absence of diabetic medication and taking a group of individuals on whom the test hadn’t been performed. This showed us that frequent HbA1c testing on diabetic patients could lead to lower rates of patient readmission. Administering a HbA1c test on patients with diabetes therefore a useful indicator of readmission rates which may prove to be valuable in the development of strategies to reduce readmission rates and costs for the care of individuals with diabetes. The data suggest further that the greater attention to diabetes reflected in HbA1c determination may improve patient outcomes and lower cost of inpatient care.

# Implementation difficulties:

* Getting accustomed to the working of the R language
* Data-set reduction
* Analyzing and removing duplicate entries in the data-set

# References:

[1] Impact of HbA1c Measurement on Hospital Readmission Rates:

Analysis of 70,000 Clinical Database Patient Records by Beata Strack,1 Jonathan P. DeShazo,2 Chris Gennings,3 Juan L. Olmo,4 Sebastian Ventura,4 Krzysztof J. Cios,1,5 and John N. Clore 6

[2] G. E. Umpierrez, S. D. Isaacs, N. Bazargan, X. You, L.M.Thaler, and A. E. Kitabchi, “Hyperglycemia: an independent marker of in-hospital mortality in patients with undiagnosed diabetes,” Journal of Clinical Endocrinology and Metabolism, vol. 87, no. 3, pp. 978–982, 2002.

[3] M. C. Lansang and G. E. Umpierrez, “Management of inpatient hyperglycemia in non critically ill patients,” Diabetes Spectrum, vol. 21, no. 4, pp. 248–255, 2008.