

Analytics by Anova in Clinical Predictions

R. Jamuna

Dept. of Computer Science, Seethalakshmi Ramaswami College, Bharathidasan University, Tiruchirappalli, India

*Corresponding Author: rjamuna2002@yahoo.co.in

DOI: <https://doi.org/10.26438/ijcse/v7i9.167170> | Available online at: www.ijcseonline.org

Accepted: 10/Sept/2019, Published: 30/Sept/2019

Abstract — Diabetes is a dreadful disorder facing the mankind from children to the older people irrespective of their age. This paper gives a mathematical model and analytics of different types of the diabetic population with different types medical managements. Diabetes is mainly a pancreatic disorder where there is insufficient secretion of insulin or improper functioning in the utilization of insulin for Glucose metabolism. Mainly there are five types of diabetes disorders such as Type-1, Type-2, Gestational diabetes, Juvenile diabetes and, MODY diabetes. In the sample population taken for study, these various types of the pancreatic disorders can be brought under control by medical managements like Allopathy, Siddha, Homeopathy and Ayurvedic treatments. The paper statistically analyses the diabetic population for number of patients under controls in various types of medical management by a mathematical model. Null hypothesis assumes that the numbers of patients under control of diabetes in various managements are independent of the type of treatments given by allopathy and other alternative methods. The powerful statistical tool ANOVA finds if there is significant difference between class means in view of variability within the separate classes. ANOVA method is used to analyze the sample population and the null hypothesis assumes that the number of patients with different level of controls is the same for different types of treatments in the sample diabetic population. The calculated value of variance ratio F is $>$ the table value at 5% level. So the null hypothesis may be rejected at 5% level of significance. It gives an inference that there is significant difference between treatments given by various methods like allopathy, siddha, ayurvedic, homeopathy in the level of control for patients. Even though diabetes can never be cured by any method but can be kept under control to avoid complications in major organs.

Keywords- MODY, STATISTICAL MODELING, Clinical Predictions

I. INTRODUCTION

Diabetes is a dreadful disorder facing the mankind from children to the older people irrespective of their age. This paper gives a mathematical model of different samples of the diabetic population. Diabetes is mainly a pancreatic disorder where there is insufficient secretion of insulin or improper functioning in the utilization of insulin for Glucose metabolism. Mainly there are five types of diabetes disorders such as Type-1, Type-2, Gestational diabetes, Juvenile diabetes and MODY diabetes taken here for the sample population. These various types of the pancreatic disorders can be brought under control by various types of medical managements like Allopathy, Siddha, Homeopathy and Ayurvedic treatments and the number of patients under control varies and tabulated below in Table [1]. Some classic symptoms of diabetes include frequent urination, extreme thirst with false hunger, gasping for breath, fatigue, blurry vision, slow healing of cuts and bruises, and sudden weight loss [1]etc. The paper analyses the diabetic population for success in various types of medical management by a mathematical model. ANOVA method is used to analyze the sample population and the results give a very significant inference that the level of control differs with the different

types of treatments in the various types of the diabetic population. Analysis of variance (ANOVA) is a very significant method which is an analysis tool used in statistics that splits the aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, but the random factors do not. Analysts use the “Analysis of the variance test” to determine the result that independent variables have on the dependent variable as a statistical model. Here the types of diabetes and types of treatment methods form the variables for the study.

II. RELATED WORKS

Diabetes Mellitus is a disorder which is affecting all people all over the world. In the recent survey of research papers in this area, “Web-based Fuzzy Expert System for Diabetes Diagnosis” by I.K. Mujawar, B.T. Jadhav, [1] presents an outline and execution of online fuzzy expert system for diabetes diagnosis. This work proposes a rule-based expert system where fuzzy logic is used. It was actualized online for the determination of diabetes disease using open source development environment. Here in this paper it is suggested that doctors, diabetes experts and patients can utilize Web-

FESDD for diabetes diagnosis as an intelligent diagnostic system. Fuzzy expert systems are able to handle imprecise data which occurs in process of disease diagnosis and treatment. In the usage of data mining techniques with the study of diabetes the paper "Diabetes Mellitus and Data Mining Techniques: A survey" by Mirza Shuja, Sonu Mittal, Majid Zaman [2] gives an outline of how data mining has found a special place in the healthcare sector. The paper also describes how these data mining techniques have been successfully applied for the prognosis of diabetes. This paper has the main focus to make a detailed survey of various data mining techniques and approaches that have been put to use for prognosis of diabetes. The research presented here is a survey focused mainly on evaluation of various computer based tools designed for prognosis of diabetes. Another research paper by K. Saravanapriya*, J. Bagyamani, "Performance Analysis of Classification Algorithms on Diabetes Dataset", [3] describes how performance analysis can be carried out using classification algorithms with diabetes data set. This paper totally analyzes the performance of the classification techniques in diabetes data set. The paper "An Overview of the Studies of Health Information Systems in Turkey" by O. Sebetci and M. Aksel [4] presents a literature review about health information systems. Health information technology gives information about health section data and data generation, compilation, analysis and gives reliable information for policy and system development. Health information technology is based on the efficiency, quality and cost of healthcare. There are positive influences of health information technologies on healthcare, but also some difficulties and failures that need more detail investigations in the future. The paper gives literature information for overview of health information systems in Turkey and other countries' systems. The results would provide a useful approach for evaluating the positive and negative factors and give ideas for the health information system users and researcher in Turkey.

III. MATERIALS AND METHODS

Diabetes is a disorder is a silent killer facing the mankind. If control of the same is not properly followed then it will damage the major organs of the body. This paper gives a mathematical model of different samples of diabetic population of all ages both male and female categories. Mainly 5 types of diabetes are identified among the sample population taken from the experimental study. They have success in control of the diabetes by various types of medical managements like Allopathy, Siddha, Homeopathy and Ayurvedic treatments. When patients have **Type 1** diabetes, the pancreas can't make insulin. This vital hormone helps your body's cells convert sugar into energy. Without it, sugar builds up in your blood and can reach dangerous levels. To avoid life-threatening complications, people with type 1 diabetes must take insulin for their entire lives. **Type-2** may be

called adult-onset diabetes. With the epidemic of obese and unhealthy overweight children, more teenagers are now developing type 2 diabetes. Type 2 diabetes was also called non-insulin-dependent diabetes. Type 2 diabetes [5] is often a milder form of diabetes than Type 1. Nevertheless, type 2 diabetes can still cause major health complications, particularly in the smallest blood vessels in the body that nourish the kidneys, nerves, and eyes. Type 2 diabetes also increases your risk of heart disease and stroke. There is no absolute cure for diabetes of any type. Type 2 diabetes can, however, be controlled with weight management, nutrition, exercise, and life style modification. Diabetes that is triggered by pregnancy is called **Gestational diabetes** (pregnancy, to some degree, leads to insulin resistance). It is often diagnosed in middle or late pregnancy. Because high blood sugar levels in a mother are circulated through the placenta to the baby, gestational diabetes must be controlled to protect the baby's growth and development. Maturity onset **diabetes** of the young" (**MODY**) refers to any of the several hereditary forms of **diabetes** mellitus caused by mutations in an autosomal dominant gene disrupting insulin production. **Juvenile diabetes** is now known as type 1 diabetes, and type-1 diabetes is most often diagnosed in children and young adults whose bodies do not make insulin. Some symptoms include frequent urination, extreme thirst, hunger (even though they are eating) and fatigue, blurry vision, slow healing of cuts and bruises, and weight loss. The paper uses the ANOVA method to analyze the sample population of around 658 patients with different types of diabetes with different number of patients under control by various treatments.

IV. METHODOLOGY

An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis. Analysis of variance (ANOVA) is an analysis [7] tool used in statistics in a sample population of 658 patients that splits the aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, but the random factors do not. The analysis of variance test is the initial step in analyzing factors that affect a given data set. Once the analysis of variance test is finished, an analyst performs additional testing on the methodical factors that measurably contribute to the data set's inconsistency. The analyst utilizes the analysis of the variance test results in an F-test to generate additional data that aligns with the proposed regression models. The test allows comparison of more than two groups at the same time to determine whether a relationship exists between them. The test analyzes multiple groups to determine the types between and within samples [6]. Analysis of variances is helpful for testing three or more variables. It is similar to multiple two-

sample t-tests. The Null hypothesis assumes that the number of patients with different level of controls is the same for different types of treatments in the sample diabetic population

V. STATISTICAL MODELING OF THE PROBLEM

Table [1]: Types of diabetes and Types of treatments

Type of diabetes \ Type of treatment	Juvenile Diabetes	Type-I diabetes	Type-II diabetes	Mody diabetes	Gestational diabetes
Allopathy	32	34	35	35	37
Siddha	34	33	36	37	35
Homeopathy	31	34	35	32	36
Aurvedic	29	26	30	28	29

Table [2]: Types of diabetes and Types of treatments: T_i

Type of Diabetes \ Type of treatment	Juvenile	Type-I	Type-II	Mody	Gestational	T_i	$\sum_{i=1}^5 (x_{ij})^2$
Allopathy	-2	0	-1	1	3	1	15
Siddha	0	-1	2	3	1	5	15
Homeopathy	-3	0	1	-2	2	-2	18
Aurvedic	-5	-8	-4	-6	-5	-28	166
$\sum_{i=1}^4 (x_{ij})^2$	38	65	22	50	39	$T = -24$	214

By statistical calculations, [8]

$$V = \sum_i \sum_j (x_{ij})^2 - \sum_i X_{ij}^2 - (T^2 / N) = 214 - ((-24)^2 / 20) = 185.2$$

$$V_2 = \sum_i \sum_j X_{ij}^2 = \sum T_i^2 / n_i = 214 - (1^2 / 5 + 5^2 / 5 + (-2)^2 / 5 + (-28)^2 / 5)$$

$$= 214 - \{814/5\} = 214 - 162.8 = 51.2$$

$$\therefore V_1 = V_2 = 185.2 - 51.2 = 134.$$

The calculated F value = 13.36

Table [2] shows the value of F for (3, 16) d.f at 5% significant level = 3.24

Therefore the calculated value of F > the table value of F at 5% level. The null hypothesis is rejected. The results are also validated by automatic calculator for validation which shows minor variation with the same prediction, tabulated below. Table [3]

Table [3]: F-Test calculation

Source of variation	d.f	Sum of squares (V_i)	Mean squares $v_i = V_i/d.f$	Variance Ratio (F)
Between batches	k-1=3	$V_1=134$	$v_1=44.67$	$F = v_1/v_2 = 13.96$
Within Batches	N-k=16	$V_2=51.2$	$v_2=3.2$	
	N-1	$V=185.2$		

Analysis of Variance Results by the calculation of one way anova calculator for validation.

F-statistic value = 15.07551

P-value = 0.00006

Between Groups Degrees of Freedom: $DF = k - 1$,

where **k** is the number of groups

Within Groups Degrees of Freedom: $DF = N - k$, where **N** is the total number of subjects

Total Degrees of Freedom: $DF = N - 1$

Table [4]: Validation Table

Data Summary for Validation				
Groups	N	Mean	Std. Dev.	Std. Error
Allopathy	5	34.6	1.8166	0.8124
Siddha	5	35	1.5811	0.7071
Homeopathy	5	33.6	2.0736	0.9274
Aurvedic	5	28.4	1.5166	0.6782
ANOVA Summary				
Source	DF(Degree of freedom)	SS(Sum of squares)	MS(Mean square)	F-Stat
Group s	3	140.2	46.7333	15.0755 0.0001
Within Group s	16	49.5992	3.1	
Total:	19	189.7992		

VI. RESULTS AND DISCUSSIONS

The null hypothesis assumes that that there is no significant difference between treatments given by various methods like allopathy, siddha, ayurvedic, homeopathy which has different control levels in the number of patients suffering from various forms of diabetes. But in the experimental sample population the various forms of diabetes treated by different medical management methods are independent of the number of patients observed to have control as their constitution and hereditary factors are different. For every individual a particular method may give control and we cannot say perfectly one method is suitable for one type of diabetes. Type-1 is almost proved to be a birth disorder by dominant hereditary factors. Other factors include lifestyle management and preventive measures. The F test proves that Null hypothesis in this sample is not holding good. But for more accurate predictions large samples have to be taken for analysis by data mining techniques.

VII. CONCLUSIONS AND FUTURE SCOPE

Therefore the calculated value of F is $>$ the table value t 5% level of significance. So the null hypothesis may be rejected It gives an inference that there is significant difference in the number of patients who have success in diabetic control by different treatments given by various medical methods like allopathy, siddha, ayurvedic and homeopathy. Each method has different impact and has different control levels in the number of patients suffering from various forms of diabetes depending upon the constitution of the individual and hereditary factors. Diabetes disorder has no absolute cure for any individual by any type of treatment. To save the humanity from complications developed in major organs control management by different treatments suited to the individual patient can be carried out. The paper can be further developed by taking large samples and studying the impact of each medical treatment method.

To conclude, diabetes disorder does require significant lifestyle [7] changes and modifications that include:

- Frequent testing of your blood sugar levels
- Careful meal planning
- Daily exercise
- Taking insulin and other medications as needed
- Avoiding stressful situations by yoga, meditation etc.

Ultimately saving the mankind from the complications of this life threatening disorder has the global vision to develop healthy Nation and disease free World.

REFERENCES

- [1]. Web-based Fuzzy Expert System for Diabetes Diagnosis I.K. Mujawar^{1*}, B.T. Jadhav², International Journal of Computer Sciences and Engineering OVol.-7, Issue-2, Feb 2019
- [2]. Mirza Shuja^{1*}, Sonu Mittal², Majid Zaman³ 1, 2 School of Computer and System Sciences, International Journal of Computer Sciences and Engineering Open Access Survey Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693.
- [3]. K. Saravanapriya* , J. Bagyamani ,”Performance Analysis of Classification Algorithms on Diabetes Dataset”International Journal of Computer Sciences and Engineering Open Access Research Paper Volume-5, Issue-9 E-ISSN: 2347-2693
- [4]. An Overview of the Studies of Health Information Systems in Turkey” by O. Sebetci and M. Aksel International Journal of Computer Sciences and Engineering Open Access Review Paper Volume-4, Issue-8.
- [5]. (www.nlm.nih.gov/medlineplus/healthtopics.html.and(www.niddk.nih.gov)
- [6]. Centers for Disease Control and Prevention. National diabetes statistics report, 2017. Centers of Disease Control and Prevention website.
- [7]. Analysis Of Variance - ANOVA [https:// www.investopedia.com/terms/a/anova.asp#ixzz54QxqcbsQ](https://www.investopedia.com/terms/a/anova.asp#ixzz54QxqcbsQ)
- [8]. www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf (PDF, 1.3 MB).

Authors Profile

Dr.Mrs.R.Jamuna is an Associate Professor working in the Department of Computer Science, Seethalakshmi Ramaswami College, Tiruchirappalli, for more than two decades. Her areas of interest include applications of Data mining, Java Programming and Computer Networks. She is a Ph.D in Computer Science from Bharathidasan University and published many research papers at the National and International level Journals. Some of her research papers are indexed in Scopus and Google scholar.

