

Research on the Influencing Factors that affecting Female Diabetes

Yuchen Li *

Department of Bioengineering, Beijing University of Chemical Technology, Beijing, China

* Corresponding Author Email: yli615@syr.com

Abstract. This paper provides an in-depth look at diabetes as a growing public health problem among women in light of globalization and lifestyle changes. Diabetes mellitus encompasses a collection of metabolic conditions marked by elevated blood glucose levels following a fasting period, typically resulting from inadequate insulin function or reduced sensitivity to insulin within the body. As per the International Diabetes Federation (IDF), the projected global population living with diabetes in 2011 was approximately 366 million. The IDF's projections indicate a potential increase to 522 million by 2030. This substantial escalation not only signifies the persistent rise in diabetes prevalence on a global scale, but also emphasizes the critical need for efficacious preventive interventions and diabetes management strategies. The prevalence of diabetes among women is not only high but rising, and there are significant gender differences in disease management and the probability of complications. The unique risks of diabetes in women include an increased risk of heart disease, stroke, kidney disease, and polycystic ovary syndrome. These complications further affect women's health and quality of life. In order to better understand the factors affecting diabetes in women, we conducted an in-depth analysis of relevant data using R language, focusing on Pima Indian women aged 21 to 81 years. The data set used in the study, the Diabetes Dataset, contained 768 observations and nine variables, all of which were Pima Indian women. This paper uses the logistic regression method to analyze these variables. The findings suggest that diabetes spectrum function is the most significant factor affecting diabetes risk in this population. This study is of great value to academic research and has far-reaching implications for formulating public health policies and optimizing clinical treatment guidelines.

Keywords: Logistic regression; female diabetes; influencing factors.

1. Introduction

With globalization and changing lifestyle patterns, diabetes mellitus has become a growing public health problem, especially among the female population. Diabetes mellitus is a state of hyperglycemia after fasting, a group of metabolic disorders characterized by hyperglycemia, usually caused by insufficient insulin action or a diminished body effect on insulin [1, 2]. As per the International Diabetes Federation (IDF), the global diabetic population was approximately 366 million in 2011. Projections from IDF indicate a potential increase to 522 million by 2030, signifying a significant upward trend in diabetes incidence. This trend challenges public health systems and underscores the critical need for diabetes prevention and management. In the female population, not only is the prevalence of diabetes high and on the rise but there are also significant gender differences in disease management and probability of complications. The unique hazards of diabetes in women include an increased risk of heart disease, stroke, kidney disease, and conditions such as polycystic ovary syndrome, complications that further exacerbate the impact of diabetes on women's health and quality of life.

There are several common types of diabetes: type I diabetes, type II diabetes, gestational diabetes, and menopausal diabetes. Type I diabetes (T1D) is primarily triggered by genetic and autoimmune factors that lead to the destruction of pancreatic beta-cells [3], whereas type II diabetes (T2D) is strongly associated with lifestyle, particularly obesity, poor dietary habits, lack of exercise, and life stress. Gestational diabetes mellitus is specific to pregnancy and is related to the weight, diet, and genetic background of pregnant women [4]. In addition, hormonal changes during menopause as we age are an essential factor contributing to diabetes in middle-aged and older women.

Type I diabetes, although primarily caused by genetics, cannot be ignored as it can lead to long-term insulin dependence and problems with glycemic control [5]. With the progress made in contemporary medical science, life expectancy for those with type 1 diabetes (T1D) remains significantly lower compared to individuals without the condition [6]. The main contributor to the higher mortality rates in this population is the prevalence of cardiovascular diseases, particularly among female patients, who show the most significant proportional increase in risk [7]. Women with T1D have twice the excess compared to men with T1D. The risk of lethal and non-lethal vascular disease is almost 40% higher [8]. Type II diabetes is usually strongly associated with weight, diet, and exercise habits, increasing the risk of complex complications such as heart disease, stroke, neuropathy, and retinopathy. Gestational diabetes, although primarily a temporary state, can increase the risk of pregnancy complications and elevate the likelihood of type II diabetes in the mother and offspring later in life [9]. In menopausal women, diabetes increases the risk of cardiovascular disease, osteoporosis, and cognitive decline.

In addition to these type-specific hazards, women with diabetes are commonly at risk for multiple organ dysfunction, including renal failure [10], diabetic foot, and other complications. Further environmental factors, stress, and sleep deprivation may also exacerbate these harms.

Research has indicated that females diagnosed with diabetes exhibit a higher susceptibility to cardiovascular complications compared to their male counterparts. The prevalence of cardiovascular disease (CVD) is reported to be 9% in men and 10% in women, with gender-specific complications including gestational diabetes mellitus and polycystic ovary syndrome (PCOS) [11]. Therefore, exploring the factors contributing to diabetes in women is valuable for academic research and has far-reaching implications for public health policy development and optimization of clinical treatment guidelines. Several studies have been conducted to reveal potential etiologic associations through regression analyses and cohort comparisons. However, these analyses are often limited by the comprehensiveness of the dataset and good study design and lack data from long-term studies of specific populations (e.g., women of different ethnic and economic backgrounds). In this context, the purpose of this paper is to conduct an in-depth analysis of relevant data in R to identify the factors that most significantly influence the risk of diabetes in women, and we will focus primarily on adult women of Pima Indian descent between the ages of 21 and 81.

2. Methodology

2.1. Data Source

The dataset I use for model analysis is called Diabetes Dataset. The Diabetes Dataset, a critical resource in diabetes research, is accumulated by the National Institute of Diabetes and Digestive and Kidney Diseases, with Mehmet Akturk acknowledged as the dataset's contributor. This collection provides essential data for understanding diabetes management and patient health trends. This dataset is credible since the provider and collector are reliable, the usability on kaggle.com is a perfect score, and there are 78.1k downloads.

2.2. Variable Introduction

There are 768 observations and nine variables in the dataset; all the observations are females from 21 to 81 years old of Pima Indian heritage. The nine variables are Pregnancies (which indicates the number of times people get pregnant), Glucose, pressure, SkinThickness (which measures triceps skin fold thickness), Insulin, BMI, DiabetesPedigreeFunction (which is a computational tool used to assess the probability of an individual developing diabetes based on their age and family history of the disease), Age, and Outcome (which indicates whether the person has diabetes, 0 represent the person does not have diabetes, 1 represent the person have diabetes).

Table 1. Different types of variables

Term	Type	Meaning
Pregnancies	univariate	the number of times people get pregnant
Glucose	univariate	triceps skin fold thickness
Pressure	univariate	
Skin Thickness	univariate	
Insulin	univariate	
BMI	univariate	
Diabetes Pedigree Function	univariate	The likelihood of diabetes was calculated based on the subject's age and family history of diabete
Age	univariate	
Outcome	univariate	

This paper uses the pairs function in R to check the distribution of all the variables (Table 1). Based on the Figure 1, most distributions of these variables are skewed to the right, such as Pregnancies, SkinThickness, Insulin, DiabetesPedigreeFunction, and Age. However, a few variable distributions are typically distributed, such as glucose, blood pressure, and BMI. Since the Outcome variable only contains 0 or 1, it is difficult to see its distribution, but this paper uses a scatter plot to check its distribution, and it shows that there are more people without diabetes.

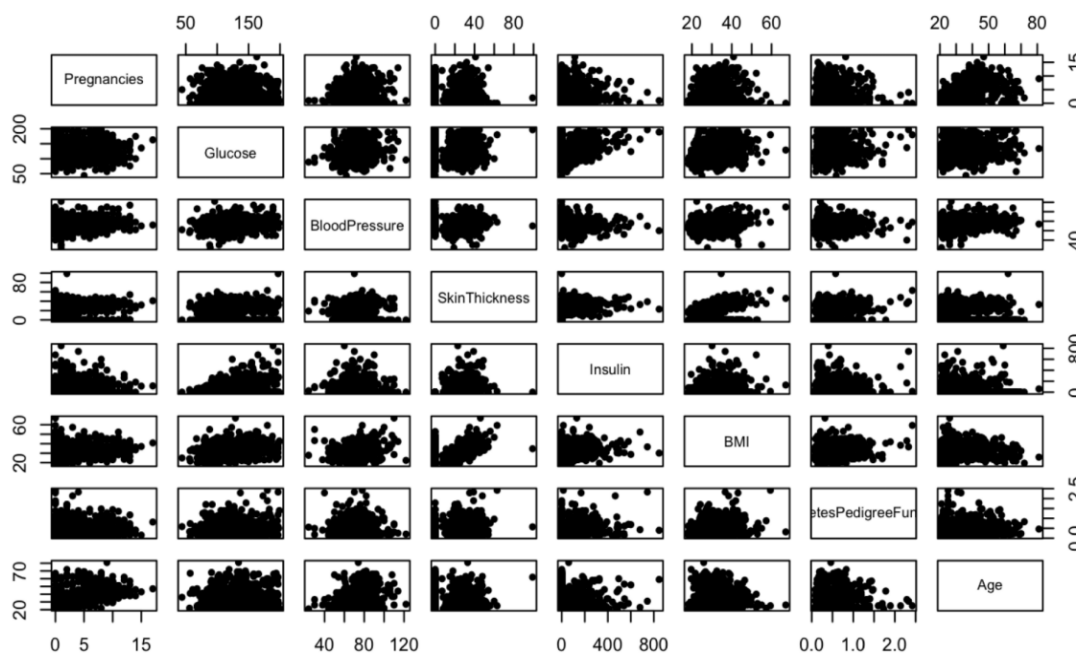


Fig. 1 Distribution of Variables

2.3. Method Introduction

In this study, the diabetes risk factors were quantitatively analyzed by the logistic regression model. The results showed that the number of pregnancies, blood glucose levels, body mass index (BMI), family history of diabetes, and age were positively correlated with the risk of diabetes, with blood glucose and BMI having the most significant effects. Specifically, each one-unit increase in the number of pregnancies, blood sugar levels, BMI, family history of diabetes, and age was associated with an increased risk of diabetes of 1.16%, 3.7%, 9.4%, 100.8%, and 16%, respectively. In contrast, for every unit increase in insulin levels and skin thickness, the risk of diabetes decreased by 0.1% and increased by 0.01%, respectively, but these associations were insignificant. These findings highlight the importance of focusing on these key risk factors when preventing and managing diabetes. In particular, controlling blood sugar and BMI should be the focus of diabetes prevention strategies. In

addition, the significant influence of family history of diabetes also suggests the critical role of genetic factors in the development of diabetes.

3. Results and Discussion

3.1. Variable Selection

There is a lot of weird data in this dataset. For example, skin thickness, body mass index, and insulin have many zeros. This study first converts all zeros to "NA" and then uses a filtering function to remove all observations with "NA" in the BMI variable. However, this study maintains that the skin thickness and insulin variables are zero, as many zeros can be seen in the skin thickness and insulin variables based on the plotted scatter plot and histogram. If you delete them all, more observations may need to be analyzed.

3.2. Logistic Regression

This study calculated the correlation coefficients between the variables to check multicollinearity and illustrated the correlation matrix using a heat map, as shown in Figure 2. The results show that there is no multicollinearity between the variables. A filter was set up for variables with a correlation coefficient greater than 0.8, but no variables were filtered out.

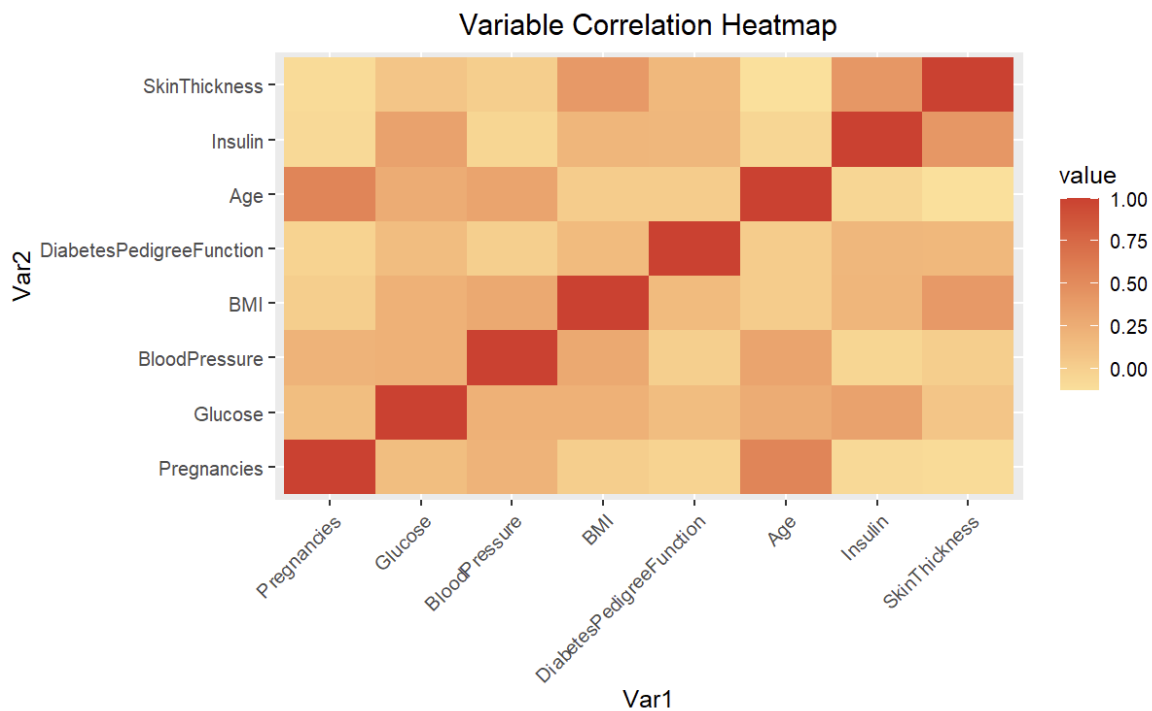


Fig. 2 Variable Correlation Heatmap

Because the dependent variables in this study were classified, Logistic regression was chosen to predict diabetes. The result after logistic regression is as follows.

The fitting results show the estimated coefficients and standard errors of the model (Table 2). In this case, the estimated coefficients showed that the number of pregnancies, blood glucose levels, BMI, family history of diabetes, and age were positively correlated with the risk of diabetes, with a significant positive correlation with *. Insulin levels and skin thickness were inversely associated with diabetes risk, neither with *. This negative correlation is not significant; that is, there is no reason to think it is a negative correlation. Only a p-value less than 0.05 can be sufficient reason to believe there is a negative or positive correlation.

Table 2. Model results.

Coefficients:	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-9.0892948	0.8004894	-11.355	< 2e-16	***
Pregnancies	0.1254415	0.0324133	3.870	0.000109	***
Glucose	0.0383267	0.0038849	9.866	< 2e-16	***
BloodPressure	-0.0090023	0.0084887	-1.061	0.288911	
SkinThickness	-0.0027742	0.0068257	-0.406	0.684421	
Insulin	-0.0013561	0.0009169	-1.479	0.139139	
BMI	0.0963914	0.0163449	5.897	3.69e-09	***
DiabetesPedigreeFunction	0.8976804	0.3000464	2.992	0.002773	**
Age	0.0107689	0.0095163	1.132	0.257791	

Note: ***, 0.001, **0.01 significance level

Here are some concrete examples of these values: The estimated coefficient showed that the risk of diabetes increased by 0.1254415 times for each additional number of pregnancies. Standard error shows the confidence interval width for the estimated coefficient. The z value shows the significance level of the estimated coefficient. In this case, the z-value is 3.870, and the p-value is 0.000109. This means that the estimated coefficient is significantly different from 0.

$$P = \frac{\exp(-9.057+0.116PRE+0.037GLU-...-0.001INS+0.0001ST)}{1-\exp(-9.057+0.116PRE+0.037GLU-...-0.001INS+0.0001ST)} \tag{1}$$

Based on the regression coefficient estimates, the following conclusions were drawn: For each additional unit of pregnancies, the probability of diabetes increases by 1.16%; for each additional unit of glucose, the probability of diabetes increases by 3.7%; or each additional unit of blood pressure, the probability of diabetes decreases by 1.1%; for each additional unit of BMI, the probability of diabetes increases by 9.4%; for each additional unit of diabetes pedigree function, the probability of diabetes increases by 100.8%; for each additional unit of age, the probability of diabetes increases by 16%; for each an additional unit of insulin, the probability of diabetes decreases by 0.1%; lastly, for each an additional unit of skin thickness, the probability of diabetes increases by 0.01%.

4. Conclusion

In this study, the diabetes risk factors were quantitatively analyzed by the logistic regression model. The results showed that the number of pregnancies, blood glucose levels, body mass index (BMI), family history of diabetes, and age were positively correlated with the risk of diabetes, with blood glucose and BMI having the most significant effects. Specifically, each one-unit increase in the number of pregnancies, blood sugar levels, BMI, family history of diabetes, and age was associated with an increased risk of diabetes of 1.16%, 3.7%, 9.4%, 100.8%, and 16%, respectively. In contrast, for every unit increase in insulin levels and skin thickness, the risk of diabetes decreased by 0.1% and increased by 0.01%, respectively, but these associations were insignificant. These findings highlight the importance of focusing on these key risk factors when preventing and managing diabetes. In particular, controlling blood sugar and BMI should be the focus of diabetes prevention strategies. In addition, the significant influence of family history of diabetes also suggests the critical role of genetic factors in the development of diabetes.

References

- [1] Alam U, Asghar O, Azmi S, et al. General aspects of diabetes mellitus. Handbook of clinical neurology, 2014, 126: 211-222.
- [2] Whiting DR, Guariguata L, Weil C et al. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract., 2011, 94: 311-321.

- [3] Blair M. Diabetes mellitus review. *Urologic nursing*, 2016, 36(1).
- [4] Xinyue Liu, Yeyi Zhu, Marissa Seamans, Roch Nianogo, Carla Janzen, Zhe Fei, Liwei Chen, Gestational diabetes mellitus and risk of neurodevelopmental disorders in young offspring: does the risk differ by race and ethnicity? *American Journal of Obstetrics & Gynecology MF*, 2023.
- [5] Bruno Vergès. Cardiovascular disease in type 1 diabetes, an underestimated danger: Epidemiological and pathophysiological data. *Atherosclerosis*, 2023.
- [6] Lind M, Svensson A M, Kosiborod M, et al. Glycemic control and excess mortality in type 1 diabetes. *N Engl J Med*, 2014, 371(21):1972–82.
- [7] Livingstone S J, Levin D, Looker H C, Lindsay R S, Wild S H, Joss N, et al. Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008–2010. *J Am Med Assoc*, 2015, 313(1): 37.
- [8] Huxley R R, Peters S A E, Mishra G D, Woodward M. Risk of all-cause mortality and vascular events in women versus men with type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*, 2015, 3(3): 198–206.
- [9] Baiju R. Shah, Denice S. Feig, Elaine Herer, Michelle A. Hladunewich, Alexander Kiss, Radha P. Kohly, Lorraine L. Lipscombe, Paul M. Yip, David Z. Cherney, Increased risk for microvascular complications among women with gestational diabetes in the third trimester. *Diabetes Research and Clinical Practice*, 2021.
- [10] Federica Piani, Isabella Melena, Kalie L, et al. Sex-related differences in diabetic kidney disease: A review on the mechanisms and potential therapeutic implications. *Journal of Diabetes and its Complications*, 2021.
- [11] Koivisto V A, et al. Cardiovascular disease and its risk factors in IDDM in Europe. *Diabetes Care*, 1996, 689–697.