Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes

Sabah Hasan Jasim

Mathematic department- College of Basic Education

Misan University

Abstract:

Diabetes is a complex group of diseases with a variety of causes. Diabetes results in abnormal levels of glucose in the blood stream for 20 sample diabetes patients. The main variables under study were ageandHbA1C are useful clinical tools regression relation was used in this study. This research include study importance of using multiple regression model to predicting the some variables like age and HbA1C that effect on diabetics with help of statistical package SPSS 22according to null hypothesis . In result show in F-value is 74,134 with 2(df) , 17degree tract show strongly significant and in T-test show T=0932 and p- value 0.365 then null hypothesis is reject and we have relation between age and diabetes T=3.115HbA1C , p- value =0.006 , T=4.046 and p-value =0.001 . Finally both age and HbA1C individually linear relation with diabetes.

Key words: Regression, Multiple Regression, Diabetes; age, HbA1C

1.1 Introduction:

Diabetes is a complex group of diseases with a variety of causes. People with diabetes have high blood glucose, also called high blood sugar or hyperglycemia diabetes is a disorder that affects the body's ability to make or use insulin. The pancreas is produce Insulin hormone that helps transport glucose (blood sugar) from the bloodstream into the cells so they can break it down and use it for fuel. Human cannot live without insulin[1].

Diabetes results in abnormal levels of glucose in the blood stream. This can cause severe short-term and long term consequences ranging from brain damage to amputations and heart disease [2]. The main variables under study were ageandHbA1C are useful clinical tools in diabetic patients. HbA1C test is called the hemoglobin A1C. The A1C test is primary test used for diabetes management and diabetes research. The HbA1C test is based on the attachment of glucose to red blood cells that

carry oxygen .The main objective of this paper is find relationship between age and HbA1C on diabetics. The data was collected from biological laboratories in misan city and the number of samples was twenty. A multiple linear regression relation is carried out to predict the values of dependent variable diabetics given a set of explanatory variables of some factors .In the next section, we give the theoretical background for the multiple linear regression after that we apply method to data by SPSS22 .

2.1-Multiple linear regression model

Regression relation was used in this study. Multiple linear regression (MLR) is a method used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is sometimes also called the predict and the independent variables the predictors. In a regression relation we study the relationship, called the regression function, between one variable Y, called the dependent variable, and several other Xi, called the independent variable. The main purpose of the multiple regression relation is to find which explanatory variables contribute to the variation of the response variable. Model equation the model expresses the value of a predict and variable as a linear function of one or more variables and an error term [3].

Model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad i=1, \ldots\ldots,n$$

Where:

Y =diabetics

X_{i1}=age

X_{i2}=HbA 1C

β_0=intercept

ε=is a random error term.

Two independent variables X_1 ,X_2.

It is up to us to decide just what independent variables we hypothesize and associated with our dependent variable of interest [12].

T- test statistic is a statistic a quantity derived from the sample used in statistical hypothesis testing. A hypothesis test is typically specified in terms of a test statistic, considered as a numerical summary of a data-set that reduces the data to one value that can be used to perform the hypothesis test. The advantage of thet-test statistic is that its sampling distribution under the null hypothesis must be calculable, either exactly or approximately, which allows p-values to be calculated. A t-test statistic shares some of the same qualities of a descriptive statistic, and many statistics can be used as both test statistics and descriptive statistics[4].

F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled[5]. The advantage of the ANOVA F-test is that we do not need to pre-specify which treatments are to be compared, and we do not need to adjust for making multiple comparisons. The disadvantage of the ANOVA F-test is that if we reject the null hypothesis, we do not know which treatments can be said to be significantly different from the others, nor [6] [7].

2.2- Null hypothesis

Usually refers to a general statement or default position that there is no relationship between two measured phenomena, or no association among groups[8].Rejecting or disproving the null hypothesis and thus concluding that there are grounds for believing that there is a relationship between two phenomena. The null hypothesis and the alternate hypothesis are terms used in statistical tests, which are formal methods of reaching conclusions or making decisions on the basis of data. The hypotheses are conjectures about a statistical model of the population, which are based on a sample of the population. The tests are core elements of statistical inference, heavily used in the interpretation of scientific experimental data, to separate scientific claims from statistical noise.

Also the statement being tested in a test of statistical significance is called the null hypothesis. The test of significance is designed to assess the strength of the evidence against the null hypothesis. Usually, the null hypothesis is a statement of no effect or no difference .It is often symbolized asH_0. The statement that is hoped or expected to be true instead of the null hypothesis is the alternative hypothesis. Symbols include H_1 and H_a [9].

2.3 -Testing of parameter's significance:

In liner regression, F-statistic is the test for relation in ANOVA approach in table (3).F- Statistic to find if the regression is significant or not if regression is not significance then Y, (diabetes) Is not dependent toX_1 ,X_2(age and HbA1C).

Consider the hypotheses as:

$H_0: B_1 = B_2 = \cdots = B_k = 0$

So model: $Y = B_0 + \varepsilon_i$

$H_{1:}$ at least one of the $B \neq 0; i = 1, \cdots n$

Model:$y_{i=}B_0 + B_1 X_{i1} + B_2 X_{i2} + \cdots + B_k X_{ik+}\varepsilon_i$ i=1………..,n

Test statistic $F = \frac{MS(refression)}{MS(residudl)}$

If F> $F_{k,n-k,\alpha}$.thenH_0 is reacted and we conclude that Y (diabetes) depend on X_1, X_2(age and HbA1C). Or H_1 is rejected and it is accepted, that means the diabetes does not depend on the X [10].

In table (3) F-value is(74.143) with 2(df) and 17 degree that show strongly significant indicating that are jointly significant.

Second if there is a Linear / relation between $X_1, X_2 and Y$ for example if increase by one unites with other all variable X_1 andX_2 age and HbA1C remain constant and diabetic Y [11].

Increase by

We check this next hypothesis

Hypothesis test about B_i:

$H_0: B_i = 0$. There is no linear relation between X_i and Y given the rest of X_i variables.

$H_0: B_i \neq 0$ There is a linear relation between X_i and Y given the rest of X_i variables.

So,

Test statistic $T = \frac{\widehat{B_i}}{std.Error(\widehat{B_i})}$

We reject H_0 if T> $t_{n-k,\alpha/2}$ 0, T< $-t_{n-k,\alpha/2}$

So we accept the null hypothesis.

$H_1: B_i$=0. No linear relationship between HbA1C and diabetes given age.

$H_1: B_i \neq 0$. There is a linear relationship given the rest observed from table (4).

 T=$-0.932$and P value 0.365 then null hypothesis is rejected and we have linear relation between age and diabetics given the rest and also when we repeat the hypothesis forgiven the rest we here T=3.115 HbA1C ,P- value = 0.006,   also according to diabetes and given the rest hypothesis for T-4.046and P-value= 0.001.Finally both age and HbA1C individually here linear relation with diabetes.

3.1-Interpretation and Discussion of results:

The relation of the data used for the purpose of research is discussed here Predictive Analytic Software (PASW) was employed for the estimation of parameter and other calculation by use SPSS. Results obtained in the study found in table (1), (2), (3) and (4) respectively and found if there is any relation between variables X$_1$,X$_2$ and Y by using model of regression of level significance is Multiple Linear Regression .

$y_{i=}B_0 + B_1X_{i1} + B_2X_{i2} + \cdots + B_kX_{ik+}\varepsilon_i$ i=1,……….,n

The regression equation for the estimation of diabetics.

$$Y = -16.466 + 15.229X_1 + 1.948X_2$$

In table (2) the coefficient of determination r^2(0.897) explained the value of dependent variable (Y) diabetes.

4.1-Conclusion:

In our results we observe that (X_1, X_2) has significant relation on diabetic's .We mention that the diabetic's(Y) are not dependent only $X_1 age, X_2 HBA1C$ but also effect on many other parameter such as diets ,weight and hormones. This study compares with similar study for diabetics patients.

Finally, we found that diabetic disease is dependent to age and HbA1C.My recommendation we will found more variables to make more complex model.

Table (1): Variables Entered for 20 sample diabetes patient: Removed (b)

Model	Variables Entered	Variables Removed	Method
1	X2, X1(b)	.	Enter
a. Dependent Variable: y			
b. All requested variables entered.			

Table (2): Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.947 (a)	.897	.885	19.02962
a. Predictors: (Constant), X2, X1				
b. Dependent Variable: y				

Table(3): Anova (b)

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	53698.398	2	26849.199	74.143	.000(b)
	Residual	6156.152	17	362.127		
	Total	59854.550	19			
a. Dependent Variable: y						
b. Predictors: (Constant), X2, X1						

Table (4):Coefficients (a)

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	-16.466	17.672		-.932	.365
	X1	15.229	3.764	.560	4.046	.001
	X2	1.948	.625	.431	3.115	.006
a. Dependent Variable: y						

Index: The value of Diabetes variables.

	X1	X2	y
1	6.00	22.00	126.00
2	6.50	35.00	140.00
3	7.00	40.00	154.00
4	7.50	35.00	169.00
5	8.00	40.00	183.00
6	8.50	41.00	197.00
7	9.00	50.00	212.00
8	9.50	55.00	226.00
9	12.00	56.00	240.00
10	11.00	60.00	298.00
11	10.00	62.00	269.00
12	9.00	50.00	240.00
13	8.00	40.00	183.00
14	7.00	45.00	154.00
15	6.00	25.00	126.00
16	5.00	15.00	97.00
17	6.00	32.00	126.00
18	5.00	33.00	140.00
19	11.00	44.00	269.00
20	10.00	39.00	184.00

References:

1-American Diabetes Association; (2007): Standards of medical care in diabetes. Diabetes Care. J. 30 Suppl. 1:S4-S41. attenuates diabetic nephropathy in rats. Clinical& Experimental Pharmacology & Physiology. 31:244-8.

2-Kumar PJ, Clark M; (2002): Textbook of Clinical Medicine. Pub: Saunders (London), pp 1099-1121,

3-Graumlich, L.J.; 1987: Precipitation variation in the Pacific Northwest (1675-1975) as reconstructed from tree rings, Annals of the Association of American Geographers 77(1), 19-29.

4- Berger, R. L., Casella, G.; (2001): Statistical Inference, Duxbury Press, Second Edition (p.374).

5- Lomax, Richard G.; (2007): Statistical Concepts: A Second Course. p. 10. ISBN 0-8058-5850-4.

6- Box, G. E. P.; (1953): "Non-Normality and Tests on Variances". Biometrika. 40(3/4): 318–335. doi:10.1093/ biomet /40.3-4.318. JSTOR 2333350.

7- Markowski, Carol A; Markowski, Edward P.; (1990): "Conditions for the Effectiveness of a Preliminary Test of Variance". The American Statistician. 44.

8-Everitt, Brian ; (1998): The Cambridge Dictionary of Statistics. Cambridge, UK New York: Cambridge University Press. ISBN 0521593468.

9-Moore, David; McCabe, George; (2003): Introduction to the Practice of Statistics (4 ed.). New York: W.H. Freeman and Co. p. 438. ISBN 9780716796572.

10- Ogunsakin.R, R.B.Ogunrinde , O.Omotoso , O.B.Adewale ; 2012: On Regression Analysis of The Relationship Between Age And Blood On Blood Pressure. International Journal Of Scientific & Technology Research Volume 1,Issue 9 , October : Issn 2277-8616.

11-Jurate B andTomasB .,2013:The Regression Model For Adults Participation In Non-Formal Education: Public vs. Private Sectors . Innovative Infotechnologies for Science, Business andEducation, ISSN 2029-1035 – 1(14) 2013 – Pp. 10-14.

12- Sabah Alsaedi, Majid Forghani; 2015: Natural Gas Consumption Regression Model for The Relationship With Population and Temperature in Misan Region (Al Amara) . Scientific journal «Herald of the Ural State University of Railway Transport» NO, 2 (26).

استخدام نموذج الانحدار المتعدد لإيجاد العلاقة بين العوامل المؤثرة للإصابة بمرض السكري

م. م. صباح حسن جاسم الساعدي

قسم الرياضيات / كلية التربية الأساسية

جامعة ميسان

الخلاصة

في هذا البحث تضمنت دراسة أهمية استخدام نموذج الانحدار المتعدد لمعرفة العلاقة لبعض المتغيرات مثل العمر والسكر التراكمي التي تؤثر على مرض السكري بمساعدة استخدام البرنامج الإحصائي SPSS 22 وقد تم استخدام فرضية العدم في تفسير النتائج الموجودة في جدول T-test و F-test حيث كانت النتائج F-value هي 134,74 ذات درجة حرية (2) و(17)وكذلك أعطت نتائج إحصائية معتمدة على جدول T-test, 939-=T عند قيمة P-value 0.365 = حيث رفض فرضية العدم واستخدام الفرضية البديلة وثم إيجاد علاقة بين العمر والسكر التراكمي عند مرض السكري حيث وجدت النتائج أن قيمة T=3.115وقيمة P تساوي 0.006 للعامل التراكمي أما بالنسبة للعمر فكانت قيمة T=4.046وقيمة P تساوي 0.01 وأخيراً يعتبر متغير العمر ومتغير السكر التراكمي للإفراد يُكون علاقة خطية مع مرض السكري .