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 age and HbA1C 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 22 according to null hypothesis. In result show in F-value is 74.134 with 2(df) , 17 degree 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=3.151 HbA1C, 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 age and HbA1C 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
Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes

Sabah Hasan Jasim

carry oxygen. The main objective of this paper is to find the 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:

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

Where:

- \( Y \) = diabetics
- \( x_{i1} \) = age
- \( x_{i2} \) = HbA 1C
- \( \beta_0 \) = intercept
- \( \varepsilon \) = 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 variable that can be used to perform the hypothesis test. The advantage of the t-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].
Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes………………Sabah Hasan Jasim

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 as $H_0$. 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 to $X_1 ,X_2$(age and HbA1C).

Consider the hypotheses as:

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

So model: $Y = B_0 + \varepsilon_i$

$H_1: \text{at least one of the } B \neq 0; i = 1,\ldots, n$

Model: $y_i = B_0 + B_1 x_{i1} + B_2 x_{i2} + \ldots + B_k x_{ik} + \varepsilon_i \quad i=1,\ldots,n$
Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes

Sabah Hasan Jasim

Test statistic $F = \frac{MS(\text{reression})}{MS(\text{residual})}$

If $F > F_{k,n-k,\alpha}$ then $H_0$ 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 $X_1, X_2 \text{and} Y$ for example if increase by one unites with other all variable $X_1 \text{and} X_2$ 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 $T = \frac{\bar{B}_i}{\text{std. Error}(\bar{B}_i)}$

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

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).

$T = -0.932 \text{ and P value} 0.365$ 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 $T = 4.046$ and 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.

$y_i = B_0 + B_1X_{i1} + B_2X_{i2} + \cdots + B_kX_{ik} + \varepsilon_i \quad i=1,\ldots,n$

The regression equation for the estimation of diabetics.
Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes

\[ Y = -16.466 + 15.229X_1 + 1.948X_2 \]

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)

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables Entered</th>
<th>Variables Removed</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X2, X1(b)</td>
<td>.</td>
<td>Enter</td>
</tr>
</tbody>
</table>

a. Dependent Variable: y
b. All requested variables entered.

Table (2): Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.947 (a)</td>
<td>.897</td>
<td>.885</td>
<td>19.02962</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), X2, X1
b. Dependent Variable: y

Table (3): Anova (b)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>2</td>
<td>26849.199</td>
<td>74.143</td>
<td>.000(b)</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>17</td>
<td>362.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>19</td>
<td>59854.550</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: y
b. Predictors: (Constant), X2, X1

Table (4): Coefficients (a)

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>T</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>-16.466</td>
<td>17.672</td>
<td>-.932</td>
</tr>
<tr>
<td></td>
<td>X1</td>
<td>15.229</td>
<td>3.764</td>
<td>.560</td>
</tr>
<tr>
<td></td>
<td>X2</td>
<td>1.948</td>
<td>.625</td>
<td>.431</td>
</tr>
</tbody>
</table>

a. Dependent Variable: y
Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes

Sabah Hasan Jasim

Index: The value of Diabetes variables.

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.00</td>
<td>22.00</td>
<td>126.00</td>
</tr>
<tr>
<td>6.50</td>
<td>35.00</td>
<td>140.00</td>
</tr>
<tr>
<td>7.00</td>
<td>40.00</td>
<td>154.00</td>
</tr>
<tr>
<td>7.50</td>
<td>35.00</td>
<td>169.00</td>
</tr>
<tr>
<td>8.00</td>
<td>40.00</td>
<td>183.00</td>
</tr>
<tr>
<td>8.50</td>
<td>41.00</td>
<td>197.00</td>
</tr>
<tr>
<td>9.00</td>
<td>50.00</td>
<td>212.00</td>
</tr>
<tr>
<td>9.50</td>
<td>55.00</td>
<td>226.00</td>
</tr>
<tr>
<td>12.00</td>
<td>56.00</td>
<td>240.00</td>
</tr>
<tr>
<td>11.00</td>
<td>60.00</td>
<td>298.00</td>
</tr>
<tr>
<td>10.00</td>
<td>62.00</td>
<td>269.00</td>
</tr>
<tr>
<td>9.00</td>
<td>50.00</td>
<td>240.00</td>
</tr>
<tr>
<td>8.00</td>
<td>40.00</td>
<td>183.00</td>
</tr>
<tr>
<td>7.00</td>
<td>45.00</td>
<td>154.00</td>
</tr>
<tr>
<td>6.00</td>
<td>25.00</td>
<td>126.00</td>
</tr>
<tr>
<td>5.00</td>
<td>15.00</td>
<td>97.00</td>
</tr>
<tr>
<td>6.00</td>
<td>32.00</td>
<td>126.00</td>
</tr>
<tr>
<td>5.00</td>
<td>33.00</td>
<td>140.00</td>
</tr>
<tr>
<td>11.00</td>
<td>44.00</td>
<td>269.00</td>
</tr>
<tr>
<td>10.00</td>
<td>39.00</td>
<td>184.00</td>
</tr>
</tbody>
</table>
Using The Model Of Multiple Regression For The Relationship Between Effective Variables In Diabetes……………..Sabah Hasan Jasim

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

الخليصة

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 22 SPSS وقد تم استخدام فرضية العدد في تفسير النتائج الموجودة في
الجدول F-value و T-test حيث كانت النتائج هي 134,74 ذات درجة حرية (2) F-value و (17)
P-value عند قيمة T= 939, T-test وكذلك أعطت نتائج إحصائية معتمدة على جدول
P-value حيث رفض فرضية العدد واستخدام الفرضية الفردية وتم إيجاد علاقة بين العمر
والسكر التراكمي عند مرض السكري حيث وجدت النتائج أن قيمة P=0.315 وقيمة T=3.115
وقيمة T=4.046 للفريق التراكمي أما بالنسبة للعمر فكانت قيمة P=0.006 وقيمة T=4.046
P=0.01 وقيمة P=0.01.

يعتبر متغير العمر ومتغير السكر التراكمي للإفراد يكون علاقة خطية مع مرض السكري.