Using Major Pathway and Compound Analysis Methods To Identify Factors Affecting Diabetes

Zainab Sami Yaseen

Southern technical university, technical institute of Basrah, Basrah, Iraq.
Emails: zainab.s.yassin@stu.edu.iq

Abstract

Legal analysis is one of the important methods to study the interrelationships between two types of variables. an important use of this analysis is to reduce the data. Many studies use this analysis as a way to study the interrelationships between two types of variables. There have been no empirical studies of the use of legal analysis as a method. From my point of view, this study aims to shed light on how to use legal analysis as a means of factor analysis, and to show how to apply it in this field by dealing with a practical problem in the active field. The applied problem includes the study of the factorial analysis, the method of the main compounds, the method of path analysis, and the compatibility between them on two types of data, represented by identifying the factors associated with diabetes, and then identifying the variables that affect the rise in the measurement of sugar two hours after eating. impact according to priority and importance.

Keywords: Principal Components Analysis; Jolliffe method; Scree Diagram method; Bartlett method; Path Analysis.

1. Introduction

When studying any of the phenomena, the expression of the relationship between the variables that control the formation of this phenomenon can be done through multiple statistical measures, some of which depend on simple methods, including the correlation coefficient, which determines the degree and nature of the relationship between the variables, despite the importance of correlation coefficients in the study However, the presence of a large number of variables affecting or involved in the formation of any phenomenon makes it difficult to interpret these coefficients efficiently and easily for two reasons, the first of which is the large number of these coefficients, and the second is that these coefficients measure the degree and type of relationship between two variables only, and this, as a result, leads to the creation of interrelationships with other variables [1,2]. Hence the need for statistical methods that enable researchers to find a solution to this problem, and the most prominent of these methods is the factor analysis. due to the interest of researchers in the field of factor analysis, several definitions have been received. We can quote several of them, which are:

- It is a statistical method that helps the researcher to study the different variables (complex phenomena) after referring them to the most important factors that affected them.
- It is a statistical method that aims to study complex phenomena to extract the most important factors that have been linked by analyzing the correlation coefficients between the variables of the phenomenon studied.
- It is one of the multivariate methods used in analyzing the correlation coefficients matrix or the covariance matrix to obtain an accurate interpretation of the relationships between the variables through the common factors behind these relationships, which are also variables, but with fewer original variables.

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

- It is also known as a branch of applied statistics that seeks a linear solution to a set of variables in a small number of factors.

A. Research Objective

This research mainly aims at how to use path analysis and principal components to reduce data (decrease data) considering several independent variables (Xs) in addition to a number of variables (Ys). We will present the mechanism of its implementation and the steps it includes, by adopting the practical application of studying the most important factors affecting diabetes.

2. Principal Components Analysis: PCA

This analysis is based on the principle of reducing a large number of explanatory variables, X's, to a smaller number of unrelated theoretical compounds called "principal compounds" while ensuring the lowest possible loss of information [3–5]. The principal components will be linear combinations of these X's, and it is very important that the units of measure are equal, otherwise we have to use the standard values to transform the matrix of X of degree (n×k), X=[xij] ; i=1, 2, …,n ; j=1, 2, …, k , to the matrix of standard values Z of degree (n×k) also, since: Z=[zij], and that:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{S_{ij}}} \qquad (1)$$

$$S_{ij} = \sum_{i=1}^{n} \left(X_{ij} - \bar{X}_j\right)^2 \ , \sum_{i=1}^{n} Z_{ij} = 0 \ , \ \sum_{j=1}^{n} Z_{ij}^2 = 1 \ , \ \sum_{j=1}^{K}\sum_{i=1}^{n} Z_{ij}^2 = K$$

Thus, the matrix Z'Z will be a matrix of simple correlations R between the explanatory variables: X1, X2, …, XK, meaning that: R = Z'Z.

The eigenvalues, denoted by the symbol λj, as well as the eigenvectors, denoted by the symbol Pj, will be extracted.

And that: a'j = [a1j a2j a3j … akj] . The values of λj are the solutions of the equation:

$$\left|R - \lambda I\right| = 0 \ , \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0 \qquad (2)$$

So that each potential value λj is associated with a potential vector aj such that it satisfies the set of equations:

$$(\mathrm{R} - \lambda_j \mathrm{I})a_j = 0 \quad ; j = 1, 2, 3, \dots, k \qquad (3)$$

With the condition: a'jaj=1. The vectors aj will then be used to re-express the principal components Fj in terms of the standard variables Zj, and in the following form:

$$F_j = a_{1j}Z_1 + a_{2j}Z_2 + \cdots . + a_{kj}Z_k = a'jZ \qquad (4)$$

The Fj corresponding to the largest latent value λj is called the first principal component, and it explains the largest proportion of the total variance in the standard data set, Zj. and that the sum of the total variance is: $\sum_{j=1}^{K} \lambda_j = K$.

For the purpose of reducing the number of combinations (dimensions) that will be used in explaining the studied phenomenon, it is not possible to use all the main compounds, but it is possible to rely on some selection rules. Remarkably, there is no general agreement or acceptance regarding this method of selection. Some researchers prefer to choose the eigenvalues that exceed the correct one to be taken into consideration. Others suggest that the selected compounds should contribute to explaining 70% or more of the total variance. In other words; The first four or five principal compounds with the

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

greatest contribution are selected with the condition: $\frac{\sum_{j=1}^{r}\lambda_j}{K} \geq 0.70$, r < k , for the first four or five compounds. Another proposal is that the accumulated percentage of the total variance be equivalent to 75% to 80% for the first five or six compounds of the main compounds.

A. Characterization of Principal Component for projection pursuit of the data

- **Projection Pursuit**

The researchers FRIEDMAN and TUKEY used the term appropriate projection, to name the technique used in the analysis of multivariate data, and it is one of the statistical methods that aims to solve the problem of diagnosing the construction of high-field data, and it does so by projecting the data on a subspace with a small field, where you search for Orthogonal projections of that data and choose the best one [6,7].

To clarify the concept of the projection vector, we assume that we have the vectors Z, V, and we want to project the vector Z onto the vector V, noting that V is a non-zero vector, and the angle between Z and V is θ, and the resulting vector from the projection of the vector Z on the vector V is the projection vector W, which can Find it through the following equation:

$$W = \|W\| \frac{V}{\|V\|} \quad (5)$$

When the direction of W is the same as that of V, but if the direction of vector W is opposite to the direction of vector V, as in the following:

$$W = -\|W\| \frac{V}{\|V\|} \quad ; \quad \|W\| = \|Z\| |\cos\theta| \quad (6)$$

For the purpose of finding orthogonal projections, we use the following law, and assuming that we have vectors (V1, V2, …, VP)) linearly independent vectors, the orthogonal projections that can be obtained are (W1, W2, …, WP) by applying the following equations:

$$h_p = V_p - \frac{V_p h_{p-1}}{h_{p-1} h_{p-1}} h_{p-1} - \dots - \frac{V_p h_1}{h_1 h_1} h_1 \quad (7)$$

Where the vectors h1, h2,…,hp are orthogonal and non-zero, and the orthogonal projections are:

$$W_1 = \frac{h_1}{\|h_1\|} \; , \quad W_2 = \frac{h_2}{\|h_2\|} \; , \; . \; . \; . \; , W_p = \frac{h_p}{\|h_p\|}$$

In 1985, GUOYING and ZHONGLIAN considered that PCA is a special case of PROJECTION PURSUIT. The first eigenvector corresponding to the first characteristic root is the best suitable projection that shows the construction of multivariate data by reducing the range of its variables. We can determine the appropriate projection for the data based on the analysis of the principal compounds, as the appropriate projection represents the first eigenvector (a_1') whose component has the largest variance ratio among the other principal components.

- **Important of PCA to Determine Projection Pursuit**

The appropriate projection for the data can be determined through PCA, as this projection has the lowest greatest possibility function under the assumption of a normal distribution of the matrix X and that the greatest possibility function for X is L(X;U,Σ,b) and that the logarithm of the greatest possibility function has a normal distribution $N(bU, b\sum b')$

which is maximized by maximizing the greatest likelihood estimators of b, Σ, U

U: vector of the averages for the explanatory variables of the population.

Σ : Matrix of covariance and covariance to a population.

$b_{1\times p}$: A projection vector belonging to R^p

$$\hat{u} = x \quad \text{and} \quad \hat{\Sigma} = \frac{n-1}{n} S$$

The maximum likelihood function for vector b is:

$L(b) = \max_{u,\Sigma} L(x, u, \Sigma, b)$ which can be found by:

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

$$f(x_1,x_2,\ldots,x_p)=\frac{1}{(2\pi)^{p/2}|\hat{\Sigma}|^{-1/2}}\exp\left[-\frac{1}{2}(x-u)\hat{\Sigma}^{-1}(x-u)\right]$$

$$Lf(x_1,x_2,\ldots,x_p)=\frac{1}{(2\pi)^{np/2}|\hat{\Sigma}|^{-n/2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-u)\hat{\Sigma}^{-1}(x_i-u)\right]$$

$$Lf(x_1,\ldots,x_p)=\frac{1}{(2\pi)^{np/2}|\hat{\Sigma}|^{-n/2}}\exp\left[-\frac{1}{2}tr\hat{\Sigma}(x_i-\bar{x})(x_i-\bar{x})\hat{\Sigma}^{-1}\right]$$

$$=\frac{1}{(2\pi)^{np/2}|\hat{\Sigma}|^{-n/2}}\exp\left[-\frac{1}{2}ntr\hat{\Sigma}\hat{\Sigma}^{-1}\right]$$

$$=\frac{1}{(2\pi)^{np/2}|\hat{\Sigma}|^{-n/2}}\exp\left[-\frac{1}{2}np\right]$$

$$LogL=Log1-Log\left[(2\pi)^{\frac{np}{2}}|\hat{\Sigma}|^{-n/2}\right]^{-1/2np}$$

$$=-\frac{np}{2}Log2\pi-\frac{n}{2}Log|\hat{\Sigma}|^{-\frac{1}{2}np}$$

$$=-\frac{np}{2}Log2\pi-\frac{n}{2}Log\frac{n-1}{n}|bSb'|^{-\frac{n}{2}p}$$

$$LogL(b)=-\frac{n}{2}(p+pLog(\frac{2\pi n}{n-1}))-\frac{n}{2}Log|bSb'|$$

$$L(b)=-\frac{n}{2}\left(p+pLog\left(\frac{2\pi n}{n-1}\right)\right)-\frac{n}{2}Log|bSb'| \quad (8)$$

So if it is $|bSb'|$ increasing, the maximum likelihood function for b is decreasing.

Assuming that the appropriate projection extracted from PCA is V, which has the lowest greatest likelihood function where $V=\max_b\{bSb'\}$

Hence, it can be said that V is equal to the first eigenvector because its component has the largest variance ratio among all the other principal components.

B. Choosing the Number of Principal Components

There are several methods used to determine the number of main compounds included in the analysis, and the most important of these methods are [8,9]:

- **Jolliffe method**

JOLLIFFE (1972) discussed four theories that were used to determine the remaining main compounds in the analysis, and these theories are (B1 BACKWARD, B1 FORWARD, B2, B4), where both theories, B1 FORWARD, B1 BACKWARD, provide for choosing a value of λ_0, and then determining the largest characteristic roots From λ_0, and according to B1 BACKWARD theorem, we start by entering the last main component corresponding to the smallest characteristic root (its value is greater than λ_0), but in the case of using B1 FORWORD, the main component corresponding to the largest characteristic root (its value is greater than λ_0) is entered. For both theories we perform PCA, and extract again the characteristic roots greater than λ_0, and we continue to repeat the PCA and find the characteristic roots greater than λ_0 until we see no value of the characteristic root greater than λ_0, at which point we stop the analysis. As for the B2 theory, it stipulates a one-time PCA procedure, where we choose the main compounds corresponding to the characteristic roots greater than λ_0, and the rest of the compounds are neglected.

After JOLLIFFE studied the previous theories, he began to criticize the two theories, B1 BACKWARD and B1 FORWARD, because these theories take a long time to apply.

He emphasized that theories B2 and B4 are suitable methods for determining the main compounds included in the analysis. In 1958, the researcher KAISER determined the value of 0 λ, and made it equal to one, but JOLLIFFE argued the value assumed by KAISER, usually that using the value 1 = 0 λ leads to the neglect of a large amount of important variables, which makes the proportion of

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

variance explained by the remaining variables in The analysis is a small percentage that does not reflect the studied phenomenon. Therefore, Jolliffe suggested that the value of 0 λ be equal to 0.7, that is, it enters the main compounds corresponding to the characteristic roots greater than 0.7, and the rest of the compounds are neglected from the analysis.

- **Scree Diagram method**

This method was suggested by CATTELL (1965) and called it SCREE because it counts the shape that results from drawing the values of the distinctive roots (vertical axis) against the distinctive root number (horizontal axis). The important distinctive roots, while the lower surface of the rock rests on the unimportant distinctive roots.
It was developed by JOLLIFFE (1986), where he used the logarithm of the characteristic roots against the characteristic root number (i) in the drawing, so the number of principal compounds to be selected is equal to (i) corresponding to an elbow in curve. The point at which this bending end is where the large characteristic roots stop and the small characteristic roots begin, the compounds of which are excluded from the analysis.

- **Bartlett method**

Bartlett developed in 1951 a formula according to which he tested the P-K equality from the characteristic roots extracted from the variance's matrix, and suggested the following formula to test that hypothesis

$$\chi^2 = (n-1)(p-k)\log(a_0/g_0) \tag{9}$$

p: the number of explanatory variables.
k : the number of the last distinct root to enter into the analysis.
a_0: the mean of all characteristic roots extracted from the correlation matrix and excluded from the analysis.
g_0: the geometric mean of all characteristic roots extracted from the correlation matrix and excluded from the analysis.
Normally, the previous statistic has the approximate distribution of χ^2, but in 1979 K.V. MARDIA discovered that the previous formula does not have the distribution of χ^2, so he corrected it by making an adjustment to the value of n.
And based on the characteristic roots extracted from the covariance and covariance matrix, apply the following formula:

$$\chi^2 = n'(p-k)\mathrm{Log}\left(\frac{a_0}{g_0}\right) \tag{10}$$

To test the hypothesis:

$$H_0 = \lambda_p = \lambda_{p-1} = \ldots = \lambda_{k+1}$$

And from the following equation we determine the calculated χ^2 value.

$$\chi^2 = n'(p-k)\mathrm{Log}\left(\frac{a_0}{g_0}\right) \tag{11}$$

$$n' = n - \left(\frac{2p+11}{6}\right) \quad \text{and} \quad a_0 = (\lambda_{k+1}+\ldots+\lambda_p)/(p-k)$$

$$g_0 = (\lambda_{k+1} \times\ldots\times \lambda_p)^{1/(p-k)}$$

If the calculated χ^2 is less than the tabulated χ^2 and with a degree of freedom ½(p-k+2)(p-k-1), we accept the hypothesis H0, that is, the main compounds corresponding to those roots are equal in terms of the ratio of total variance for all compounds, and therefore it is neglected.

3. Path Analysis

Path analysis is one of the methods used in analyzing the relationship that measures the degree of interdependence and commitment between variables and then analyzing it into direct and indirect effects. Path analysis relies mainly on analyzing the relationships between variables in causal models based on scientific theories or based on logical foundations. However, this does not mean that the researcher works to prove cause and effect between the variables in the causal model, just as the existence of a relationship between two variables does not mean that The independent variable is a cause of the dependent variable or the dependent variable is a consequence of the independent variable [10,11].

Path analysis that studies causal models does not, in fact, deviate from this logic, as there is no attempt in path analysis to prove the existence of a cause or effect among the variables, but this does not prevent the researcher from thinking causally, as he says (Blalock 1961): "Causal thinking belongs in

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

the form of Completely to theoretical levels where causal laws cannot be translated experimentally, but that does not prevent the researcher from thinking causally and building causal models that enable him to understand the relationships between variables, so that these models can be tested indirectly. (Wright) says: "We do not aim to analyze the paths to derive a causal relationship between a group of variables using the values of correlation coefficients, but rather we aim to apply this method of data analysis methods to a causal model that we assume on a specific theoretical basis."

This analysis has the possibility of knowing the truth about the relationships between the variables, whether they are causal or pseudo, as this is shown by dividing the correlation coefficient into its primary components in terms of showing the irrational effect of the variables and knowing the ratio of the pseudo effects to the total effects (direct and indirect).This method has the basic assumptions of the ordinary least square's method (O.L.S) and differs from it by using correlations instead of using covariances. Therefore, all the estimated coefficients are the standard regression coefficients [12,13].

If we want to use this method, we must verify the assumptions of path analysis, which are:

1) The relationship between the variables should be linear.
2) The relationship between the variables should be additive, i.e. there is no interaction between the variables.
3) The residual variables are not related to each other or to other variables in the model assumed by the researcher, i.e. it is assumed that the correlation coefficients between the residuals and all external variables are equal to zero.
4) The causal relationships in the model are unidirectional and the inverse (reciprocal) relationships between the variables are excluded.
5) The time sequence must be taken into account when arranging the variables, as they are placed in a sequential manner according to the causal priority.
6) All the changes in the model are numerical.

A. General concepts on some terms that will be used in path analysis [14,15]

- **Endogenous variable:** It is the variable whose differences are determined by causes (variables) located outside or within the causal model.
- **Exogenous variable:** It is the variable whose differences are determined by variables outside the scope of the causal mathematical model. Therefore, there is no attempt to identify or clarify the external variables and their relationships with the rest of the variables in the model.
- The internal variable is treated as a dependent variable for a group of variables and appears in another group of variables as an explanatory variable (independent), as the model includes internal and external variables at the same time.
- **Measured Variable:** Measured variables include the internal and external variables that appear in the causal model. Random error is an unmeasured variable appended to the internal variable in the path diagram to account for changes not explained by the external variables and is commonly called the latent variable.
- **Direct Effect:** These effects are part of the total effects, and they express the direct effect of a group of causal variables on the affected variable when the intermediate variables are constant.
- **In Direct effect:** These effects are part of the total effects, and they express the effect of causal variables on the variable affected by them through a group of intermediate variables.
- **Path Coefficient:** It can be called the standard regression coefficient or the standard regression weights and reads beta (β Weights) and indicates the direct impact of the external variable j as a cause on the internal variable i as the result, that is, the path coefficient expresses the expected effect on a variable that results from the change in the standard deviation of another variable As much as unity after installing all other variables.

The path coefficient is usually denoted by the letter p, and two lowercase letters or two numbers are placed, the first of which indicates the internal variable (the result), and the second indicates the external variable (the reason), so it is written P_{ij}.

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

B. Applied part

In this chapter, factor analysis, the principal components method and the path analysis method were applied and their matching was done on the data by identifying the factors associated with diabetes and then identifying the variables that affect the increase in blood sugar levels.

- **Population frame and data collection method**

A community framework was defined for diabetic patients attending Al-Kindi Hospital / Specialized Center for Endocrinology and Diabetes.

The data was collected by adopting personal interviews of patients attending the aforementioned center, and recording this in a form for collecting information about patients and containing the following variables: Y: What is the measurement of sugar two hours after eating. X_1: sex. X_2: Age. X_3: Body mass index (equivalent to dividing the weight by the square of the height in meters). X_4: Profession. X_5: Marital status X_6: How many years has it been since you were diagnosed with diabetes? X_7: What was your blood sugar level when you were diagnosed with diabetes for the first time? X_8: Does any of the family members have diabetes? X_9: How would you describe your state of health? X_{10}: Do you see a doctor to monitor your diabetes? X_{11}: What do you use to control diabetes. X_{12}: Do you perform certain exercises after having diabetes? X_{13}: What is the most difficult diabetes treatment for you to follow? X_{14}: Do you measure sugar in the following times? X_{15}: Do you measure sugar before or after food? X_{16}: How long do you keep on the diet? X_{17}: Have your eating habits changed after you had diabetes? X_{18}: Do you suffer from numbness and pain in the lower extremities? X_{19}: The kind of bread you eat. X_{20}: Do you get up at night to urinate more than three times? X_{21}: Are you thirsty? X_{22}: Do you suffer from poor appetite? X_{23}: Do you suffer from weight loss or gain since diabetes? X_{24}: Have you had hypoglycemia? X_{25}: Were there any diseases in your eyes during your diabetes? X_{26}: Do you suffer from vascular sclerosis? X_{27}: Do you suffer from weak sexual ability? X_{28}: Do you have a stroke? X_{29}: Do you suffer from excess fat? X_{30}: Did you enter the hospital because of X_{31}: How many times have you been hospitalized? X_{32}: Has a doctor told you that you have high blood pressure? X_{33}: Do you have problems with your kidneys after diabetes? X_{34}: Do you suffer from amputation? X_{35}: What is the ideal percentage of sugar before food? X_{36}: What is the ideal percentage of sugar two hours after eating? X_{37}: Have the signs of diabetes been clarified for you, in case of increase (frequent urination and dry mouth) or decrease (palpitations, tremor, sweating, loss of concentration, fainting) by doctors.

Patients with diabetes were chosen completely randomly. Each patient attending the specialized center for endocrinology and diabetes was within the community framework during the period of data collection from (4/12/2023) to (11/1/2024) without leaving a specific group of females. Or males, the study included a sample of 315 patients.

4. The results of the principal components analysis and their interpretation

Table 1: Communalities of the explanatory variables in the main compounds

variables	Communalities	variables	Communalities
x1	.890	x17	.867
x2	.757	x18	.665
x3	.600	x19	.555
x4	.515	x20	.571
x5	.741	x21	.716
x6	.723	x22	.442
x7	.462	x23	.612
x8	.530	x24	.656
x9	.685	x25	.486
x10	.568	x26	.532
x11	.640	x27	.901
x12	.525	x28	.533
x13	.539	x29	.676
x14	.702	x30	.868
x15	.605	x31	.832
x16	.845	x32	.636

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

		x33	.664

Table 2: The values of the characteristic roots and the variance ratios for each compound.

Main components	Polding variance	Principal Component Variation %	characteristic roots
1	12.959	12.959	4.795
2	20.822	7.863	2.909
3	27.331	6.510	2.409
4	33.063	5.732	2.121
5	37.507	4.443	1.644
6	41.805	4.298	1.590
7	45.906	4.101	1.517
8	49.591	3.685	1.363
9	53.078	3.487	1.290
10	56.448	3.370	1.247
11	59.595	3.147	1.164
12	62.533	2.938	1.087
13	65.368	2.835	1.049

After analyzing the simple correlation matrix, we were able to distinguish thirteen of the main components (depending on taking the characteristic roots greater than one) that explain (65.368%) of the total variance, and the variables affecting each main component were distinguished through the characteristic vector matrix (Matrix of Rounded Factors) These values are chosen when they are greater than or equal to 0.5 in order to be considered as having a significant effect, as follows:

- **The first factor:** The first factor reflected great importance by explaining (12.959%) of the total variance. This factor included two variables that had a significant impact, namely the thirtieth variable (did you enter the hospital because of) and the thirty-first variable (what is the number of times you were hospitalized) and this factor can be called the hospitalization factor.
- **The second factor:** The second factor explained (7.863%) of the total variance. This factor included three variables, two variables that had a significant impact. They are the sixth variable (how many years have passed since you discovered diabetes) and the ninth variable (how would you describe your health condition), and the other variable had an opposite significant effect, which is the variable Eighteenth (Do you suffer from numbness and pain in the lower extremities) This factor can be called the health condition factor.
- **The third factor:** The third factor explained (6.510%) of the total variance, and this factor included four variables, two variables with a significant impact, the fourth variable (occupation) and the fifth variable (marital status). Have your diabetes signs been clarified (in case of increase or decrease) and this factor can be called the psychological state factor.
- **The Fourth factor:** The fourth factor explained (5.732%) of the total variance. This factor included three variables with a significant impact. The first variable (sex) and the twelfth variable (do you perform certain exercises after you have diabetes) and the twenty-seventh variable (do you suffer from poor sexual ability This factor can be called the gender factor.
- **The Fifth factor:** The fifth factor explained (4.443%) of the total variance, and this factor included two variables that had a significant impact, namely the sixteenth variable (how long do you continue with the diet) and the seventeenth variable (did your eating habits change after you had diabetes) and this factor can be called a factor diet.
- **The sixth factor:** The sixth factor explained (4.298%) of the total variance, and this factor included two variables that had a significant impact: the thirty-fifth variable (what is the ideal sugar level before food) and the thirty-sixth variable (what is the ideal sugar percentage two hours after eating) and this factor can be called the sugar percentage factor idealism.
- **The seventh factor:** The sixth factor explained (4.101%) of the total variance, and this factor included three variables that had a significant impact: the seventh variable (what was the blood sugar level when you were diagnosed with diabetes for the first time) and the thirty-second variable (did a doctor tell you that you had blood pressure) And the thirty-third variable (Do you have kidney problems after diabetes) and this factor can be called the factor of diseases associated with diabetes.

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

- **The eighth factor:** The eighth factor explained (3.487%) of the total variance, and this factor included two variables, one of which has a significant effect is the fourteenth variable (do you measure sugar in the following times) and the other variable with an inverse significant effect is the fifteenth variable (do you measure sugar before food or after it) and this factor can be called the sugar measurement factor.
- **The ninth factor:** The ninth factor explained (3.370%) of the total variance, and this factor included two variables that had a significant impact, namely the twentieth variable (do you get up at night to urinate more than three times) and the twenty-first variable (do you suffer from thirst) and this factor can be named Treating symptoms of high blood sugar.
- **The tenth factor:** The tenth factor explained (3.685%) of the total variance, and this factor included three variables with a significant impact, namely the twenty-sixth variable (do you suffer from vascular sclerosis), the twenty-eighth variable (do you have a stroke) and the thirty-fourth variable (do you suffer from Amputation) This factor can be called the factor of arterial disease.
- **The eleventh factor:** The eleventh factor explained (3.147%) of the total variance, and this factor included two variables, one of which had a significant effect, the twenty-fourth variable (did you have hypoglycemia) and the other variable had an opposite significant effect, which was the eleventh variable (what do you use to control diabetes) and can Label this factor with the type of treatment used.
- **The twelfth factor:** The twelfth factor explained (2.938%) of the total variance, and this factor included one variable with a significant effect, which is the twenty-ninth variable (Do you suffer from an increase in fats) and this factor can be called the factor of increased fats.
- **The thirteenth factor:** The thirteenth factor explained (2.835%) of the total variance, and this factor included one variable with a significant impact, which is the twenty-third variable (Do you suffer from weight gain or loss since diabetes) and this factor can be called the weight factor.

A. Path Model

Depending on the logical relationships as well as what was determined by the medical theory about diabetes, and for the purpose of determining the relationships between the variables, the following equations were reconciled:

$$X_{16} = P_{1613}X_{13} + P_{1617}X_{17} + P_{1619}X_{19} + P_{1631}X_{31} + P_{16u16}u_{16} \tag{12}$$

$$X_{18} = P_{186}X_6 + P_{189}X_9 + P_{1814}X_{14} + P_{1817}X_{17} + P_{18u18}u_{18} \tag{13}$$

$$X_{21} = P_{2114}X_{14} + P_{2117}X_{17} + P_{2120}X_{20} + P_{2137}X_{37} + P_{21u21}u_{21} \tag{14}$$

$$X_{11} = P_{112}X_2 + P_{113}X_3 + P_{116}X_6 + P_{1112}X_{12} + P_{1119}X_{19} + P_{1120}X_{20} + P_{1123}X_{23} + P_{1130}X_{30} + P_{11u11}u_{11} \tag{15}$$

$$= P_{Y1}X_1 + P_{Y2}X_2 + P_{Y3}X_3 + P_{Y4}X_4 + P_{Y5}X_5 + P_{Y6}X_6 + P_{Y7}X_7 + P_{Y8}X_8 + P_{Y9}X_9 + P_{Y10}X_{10} + P_{Y11}X_{11} + P_{Y12}X_{12} + P_{Y13}X_{13} + P_{Y14}X_{14} + P_{Y15}X_{15} + P_{Y16}X_{16} + P_{Y17}X_{17} + P_{Y18}X_{18} + P_{Y19}X_{19} + P_{Y20}X_{20} + P_{Y21}X_{21} + P_{Y22}X_{22} + P_{Y23}X_{23} + P_{Y24}X_{24} + P_{Y25}X_{25} + P_{Y26}X_{26} + P_{Y27}X_{27} + P_{Y28}X_{28} + P_{Y29}X_{29} + P_{Y30}X_{30} + P_{Y31}X_{31} + P_{Y32}X_{32} + P_{Y33}X_{33} + P_{Y34}X_{34} + P_{Y35}X_{35} + P_{Y36}X_{36} + P_{Y37}X_{37} + P_{Yuy}u_y \tag{16}$$

If we study the determination cases for each equation in the supposed causal model, we get a one-time Just Identified Model, because the number of correlations between variables, for example in equations ((1), (2), (3)) equals $\left[\frac{n(n-1)}{2} = 10\right]$ (n represents the number of variables measured) which equals the number of path coefficients (4) in addition to the perfect external correlations between the variables (6).

As for equation (4), the number of correlations is equal to $\left[\frac{9(9-1)}{2} = 36\right]$ and is equal to the number of path coefficients (8) in addition to the complete external correlations between the variables (28).

Finally, the equation for the internal variable Y shows that the number of correlations for this equation is equal to $\left[\frac{38(38-1)}{2} = 703\right]$ and the number of path coefficients is equal to (37) in addition to the complete external correlations between the variables (6).

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

After studying the specific cases for each equation, we used the method of path analysis using the method of the greatest possibility (Maximum Likelihood) to extract the coefficients of the path using the program (Amos21), and the following results were reached:

$$X_{16} = 0.0975X_{13} + 0.7913X_{17} - 0.0349X_{19} - 0.0732X_{31} + 0.2434u_{16} \quad (17)$$

$$X_{18} = -0.2273X_6 - 0.4078X_9 - 0.1723X_{14} + 0.0936X_{17} + 0.1278u_{18} \quad (18)$$

$$X_{21} = .0244X_{14} + 0.1071X_{17} + 0.4140X_{20} - 0.1197X_{37} + 0.1874u_{21} \quad (19)$$

$$X_{11} = -0.2895X_2 - 0.2550X_3 + 0.3161X_6 + 0.1114X_{12} - 0.1510X_{19} + .0529X_{20} - 0.1397X_{23} - 0.2550X_{30} + 0.2254u_{11} \quad (20)$$

$$Y = -0.2023X_1 - 0.0382X_2 - 0.0536X_3 - 0.0211X_4 + 0.1320X_5 + 0.2595X_6 - 0.0437X_7 - 0.0387X_8 + 0.2065X_9 - 0.0016X_{10} - 0.0104X_{11} + 0.0049X_{12} - 0.0297X_{13} + 0.0149X_{14} + 0.0702X_{15} + 0.1545X_{16} + 0.1071X_{17} - 0.1367X_{18} - 0.0516X_{19} - 0.0475X_{20} - 0.1564X_{21} - 0.0698X_{22} - 0.0405X_{23} + 0.1130X_{24} + 0.0814X_{25} - 0.0320X_{26} + 0.1067X_{27} - 0.1076X_{28} + 0.0906X_{29} - 0.0751X_{30} + 0.0333X_{31} - 0.1045X_{32} + 0.0115X_{33} + 0.1333X_{34} - 0.0558X_{35} + 0.0943X_{36} - 0.0215X_{37} + 0.5748u_y \quad (21)$$

5. Interpret the results of the path analysis

Depending on the Amos 21 statistical program, the following path analysis results were reached:

A. The following internal variable equation X_{16}

$$X_{16} = 0.0975X_{13} + 0.7913X_{17} - 0.0349X_{19} - 0.0732X_{31} + 0.2434u_{16} \quad (22)$$

This equation refers to the direct effects of the external variables $X_{13}, X_{17}, X_{19}, X_{31}$on the internal variable X_{16}, and because the path coefficients show the relative importance, the effect of these variables according to the relative importance is as follows: - X_{17}, X_{13} and this means that the change in eating habits after diabetes is X_{17} It comes in the first place and has a significant direct effect of 0.7913, which shows that changing dietary habits works to increase the duration of the diet for the patient.

Followed by the effect of the variable X_{13}, which has a significant direct effect of 0.0975, as 59% of patients consider diet as the most difficult item in the treatment of diabetes. It was shown that the variable type of bread X_{19} and the variable number of times you were hospitalized X_{31} had no direct significant effect on the variable X_{16}.

The tests for the significance of the path coefficients are shown in Table 3.

Table 3: Significant tests of the variable equation path coefficients X_{16}.

			Estimate	S.E.	C.R.	P
x16	**<---**	**x13**	**.1474**	**.0579**	**2.5444**	**.0109**
x16	**<---**	**x19**	**-.0881**	**.0949**	**-.9284**	**.3532**
x16	**<---**	**x31**	**-.0669**	**.0350**	**-1.9102**	**.0561**
x16	**<---**	**x17**	**1.4922**	**.0715**	**20.8557**	*******

We note that the P-value is less than ($\alpha = 0.05$), and therefore the effect of the two variables X_{13} and X_{17} has a significant effect on the variable X_{16} and the effect of X_{19} and X_{31} is insignificant (less than 0.001).

Table 4: Direct effects on the internal variable X_{16}

internal variable	external variables	direct or overall effects	Simple Correlations
X_{16}	X_{17}	0.7913	0.799
	X_{13}	0.0975	0.208

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

It is clear from table 4 that the difference between the simple correlation and the total correlation for each variable results from the existence of bilateral correlations between the external variables. and the value of R^2 for the equation of the variable X_{16} is equal to (0.68).

B. The following internal variable equation X_{18}

$$X_{18} = -0.2273X_6 - 0.4078X_9 - 0.1723X_{14} + 0.0936X_{17} + 0.1278u_{18} \qquad (23)$$

This equation refers to the direct effects of the external variables X_6, X_9, X_{14}, X_{17} on the internal variable X_{18}, and because the path coefficients show the relative importance, the effect of these variables according to the relative importance is as follows: - X_6, X_9, X_{14}, which means X_{14} comes in the first place and has an effect Direct and inverse significant of 0.1723, which shows that some patients every two months measure sugar in the Endocrine and Diabetes Center and do not measure it at home, and therefore they will not feel the rise in sugar, and this rise will lead to damage to the lower nerves and cause a feeling of numbness and pain in the lower extremities. Followed by the effect of the variable X_6, which has a direct, inverse, significant effect of 0.2273, as the state of pain and numbness in the lower extremities usually occurs after ten years or more, and we note that 67%) of the study sample suffer from numbness and pain in the lower extremities for a period of less than ten years, and the variable X_9 has A direct negative significant effect of 0.4078, as the majority of patients when you ask them about their health status, the answer is weak at a rate of (72%), which affects the psychological state of the patient and leads to high sugar, which causes peripheral nerve damage. It has a significant direct effect on the variable X_{18}. The tests for the significance of the path coefficients are shown in Table 5.

Table 5: Significant tests for the variable equation path coefficients X_{18}

			Estimate	S.E.	C.R.	P	Label
x18	<---	**x17**	**.0882**	**.0522**	**1.6902**	**.0910**	
x18	<---	**x14**	**-.0883**	**.0285**	**-3.0963**	**.0020**	
x18	<---	**x9**	**-.3974**	**.0557**	**-7.1405**	*******	
x18	<---	**x6**	**-.1153**	**.0286**	**-4.0257**	*******	

We note that the value of ((P-value is less than $\alpha = 0.05$), and for this reason, the variables have a significant effect on the variable X_{18}, and the variable X_{17} has a non-significant effect.

Table 6: Direct effects on the internal variable X_{18}

internal variable	external variables	direct or overall effects	Simple Correlations
X_{18}	x_6	-0.2273	-0.357
	X_9	-0.4078	-0.501
	X_{14}	-0.1723	-0.212

It is clear from the aforementioned table that the difference between the simple correlation and the total correlation for each variable results from the presence of bilateral correlations between the external variables. and the value of R^2 for the equation of the variable X_{18} is equal to (0.33).

C. The following internal variable equation X_{21}:

$$X_{21} = .0244X_{14} + 0.1071X_{17} + 0.4140X_{20} - 0.1197X_{37} + 0.1874u_{21} \qquad (24)$$

This equation refers to the direct effects of the external variables X_{14}, X_{17}, X_{20}, X_{37} on the internal variable X_{21}, and since the path coefficients show the relative importance, the effect of these variables according to the relative importance is as follows: - X_{20}, X_{37}, and this means that the case of frequent urination at night comes X_{20}. In the first place, it has a direct and significant effect of 0.4140, and it shows that high sugar increases the patient's urine and increases his thirst. followed by the effect of

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

the variable X_{37}, which has a direct, inverse, significant effect of 0.1197, as some doctors do not inform patients of the symptoms of high and low diabetes, so the patient does not know that the cause of his thirst is high sugar. A significant direct effect on the variable X_{21}.

The tests for the significance of the path coefficients are shown in Table 7.

Table 7: Significant tests of the variable equation path coefficients X_{21}

	Estimate	S.E.	C.R.	P
x21 <--- x20	.4095	.0590	6.9384	***
x21 <--- x37	-.1180	.0593	-1.9912	.0465
x21 <--- x17	.1114	.0630	1.7692	.0769
x21 <--- x14	.0138	.0346	.3987	.6901

We note that the P-value is less than $\alpha = 0.05$. Therefore, the variables X_{20} and X_{37} have a significant effect on the variable X_{21}, and the two variables X_{14} and X_{17} have a significant effect on the variable X_{17}.

Table 8: Direct effects on the internal variable X_{21}

Internal variable	external variables	direct or overall effects	Simple Correlations
X_{21}	X_{20}	0.4140	0.415
	X_{37}	-0.1197	-0.126

It is clear from the table 8, that the difference between the simple correlation and the total correlation for each variable results from the existence of bilateral correlations between the external variables. and the value of R^2 for the equation of the variable X_{21} is equal to (0.20).

D. The following internal variable equation X_{11}

$$X_{11} = -0.2895X_2 - 0.0697X_3 + 0.3161X_6 + 0.1114X_{12} - 0.1510X_{19} + 0.0529X_{20} - 0.1397X_{23} - 0.2550X_{30} + 0.2254u_{11} \quad (25)$$

This equation refers to the direct effects of the external variables X_2, X_3, X_6, X_{12}, X_{19}, X_{20}, X_{23}, X_{30} on the internal variable X_{11}, and since the path coefficients show the relative importance, the effect of these variables according to the relative importance is as follows: - X_6, X_{12}, X_{19}, X_{23} , X_{30}, X_2 This means that the variable X_6, comes in the first place and has a direct significant effect of 0.3161 and shows that over the years the patient is converted to insulin use.

Followed by the effect of the variable X_{12}, which has a direct significant effect of 0.1114, as exercise reduces blood sugar, then followed by the effect of the variable X_{19}, which has a direct inverse significant effect of 0.1510, since 86% of the study sample use natural bread, which is one of the causes of high sugar. And the treatment changed from pills to insulin or increasing the dose of insulin after a period of time, then followed by the effect of the variable X_{23} and it has a direct inverse significant effect of 0.1397, as insulin works to open the appetite and thus increase weight, then followed by the effect of the variable X_{30} and it has a direct inverse significant effect of 0.2550 As some patients do not adhere to the diet and do not measure the blood sugar level after taking an insulin dose, and thus suffer from high blood sugar, which necessitates admission to the hospital, and finally the effect of the variable X_2 and has a direct inverse significant effect of 0.2895, as diabetes affects all age groups, and it has been shown The body mass index variable X_3 and the variable getting up at night in order to urinate more than three times X_{20} have no direct significant effect on the variable X_{11} . the tests for the significance of the path coefficients are shown in Table 9.

Table 9: Significant tests of the variable equation path coefficients X_{11}

	Estimate	S.E.	C.R.	P
x11 <--- x3	-.0390	.0352	-1.1069	.2683
x11 <--- x19	-.2519	.0914	-2.7551	.0059

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

			Estimate	S.E.	C.R.	P
x11	<---	x12	.0843	.0429	1.9630	.0496
x11	<---	x2	-.1948	.0490	-3.9750	***
x11	<---	x23	-.1066	.0421	-2.5309	.0114
x11	<---	x20	.0626	.0662	.9466	.3439
x11	<---	x6	.2121	.0417	5.0878	***
x11	<---	x30	-.1432	.0338	-4.2327	***

We note that the value of (P-value) is less than $\alpha = 0.05$ and therefore the effect of the variables is significant in the variable X_{11} and not significant in the variables X_3 and X_{20} .

Table10: Direct effects on the internal variable X_{11}

Internal variable	external variables	direct or overall effects	Simple Correlations
X11	X2	-0.2895	-0.352
	X6	0.3161	0.138
	X12	0.1114	0.194
	X19	-0.1510	-0.160
	X23	-0.1397	-0.115
	X30	-0.2550	-0.415

It is clear from the aforementioned table that the difference between the simple correlation and the total correlation for each variable results from the existence of bilateral correlations between the external variables. and the value of R^2 for the equation of the variable X_{11} is equal to (0.33).

E. The following internal variable Y equation

$$\begin{aligned} Y = {} & -0.2023X_1 - 0.0382X_2 - 0.0536X_3 - 0.0211X_4 + 0.1320X_5 + 0.2595X_6 - 0.0437X_7 \\ & - 0.0387X_8 + 0.2065X_9 - 0.0016X_{10} - 0.0104X_{11} + 0.0049X_{12} - 0.0297X_{13} \\ & + 0.0149X_{14} + 0.0702X_{15} + 0.1545X_{16} + 0.1071X_{17} - 0.1367X_{18} \\ & - 0.0516X_{19} - 0.0475X_{20} - 0.1564X_{21} - 0.0698X_{22} - 0.0405X_{23} \\ & + 0.1130X_{24} + 0.0814X_{25} - 0.0320X_{26} + 0.1067X_{27} - 0.1076X_{28} \\ & + 0.0906X_{29} - 0.0751X_{30} + 0.0333X_{31} - 0.1045X_{32} + 0.0115X_{33} \\ & + 0.1333X_{34} - 0.0558X_{35} + 0.0943X_{36} - 0.0215X_{37} \\ & + 0.5748u_y \end{aligned} \quad (26)$$

This equation refers to the direct effects of the external variables on the internal variable Y, and since the path coefficients show the relative importance, the effect of these variables according to the relative importance is as follows:- X_{21}, X_{18}, X_{24}, X_{34}, X_9 , X_6, This means that the variable X_6 comes in the first place and has a significant direct effect of 0.2595, indicating that over the years, a person cannot control the sugar level.

Followed by the variable X_9, which has a direct, significant effect of 0.2065, since 72% of the study sample had a weak answer about their health condition. The explanation for this is that diabetes is closely linked to the patient's psychological state, and therefore the patient's psychological state causes high blood sugar, followed by the variable X_{34}. and has a significant direct effect of 0.1333, showing that a state of amputation occurs when the patient is not in control of the sugar level, followed by the variable X_{24}, and has a direct significant effect of 0.1130, showing that a state of hypoglycemia may occur due to malnutrition after using an insulin syringe. The percentage of sugar before food is ideal. Therefore, insulin users are advised to measure the percentage of sugar before and after food, followed by the variable X_{18} , which has a direct, inverse, significant effect of 0.1367, showing that cases of numbness and pain in the lower extremities occur after a short period of diabetes due to non-compliance with the diet, then Followed by the variable X_{21}, which has a direct and inverse significant effect of 0.1564, which indicates that it is not required that high sugar be related to thirst at times. It was found that there are indirect effects of the variables X_{14}, X_9 , X_6 on the internal variable X_{18}, as well as indirect effects of the variables X_{37}, X_{20} on the internal variable X_{21} as in the table 12.

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

The tests for the significance of the path coefficients are shown in Table 11.

Table 11: Significant tests of the variable equation path coefficients Y

			Estimate	S.E.	C.R.	P
y	<---	x1	-.3933	.2710	-1.4514	.1467
y	<---	x4	-.0206	.0481	-.4275	.6690
y	<---	x5	.1395	.0973	1.4343	.1515
y	<---	x7	-.0318	.0438	-.7250	.4685
y	<---	x8	-.0712	.1112	-.6404	.5219
y	<---	x10	.0001	.1111	.0012	.9991
y	<---	x15	.0770	.0682	1.1288	.2590
y	<---	x37	-.0437	.1123	-.3892	.6971
y	<---	x20	-.0888	.1229	-.7227	.4699
y	<---	x17	.2353	.1959	1.2012	.2297
y	<---	x9	.4206	.1493	2.8179	.0048
y	<---	x6	.2811	.0796	3.5293	***
y	<---	x13	-.0443	.1018	-.4353	.6633
y	<---	x31	.0517	.1340	.3861	.6994
y	<---	x19	-.1191	.1608	-.7409	.4588
y	<---	x3	-.0460	.0630	-.7308	.4649
y	<---	x12	.0003	.0783	.0040	.9968
y	<---	x30	-.0885	.1293	-.6843	.4938
y	<---	x23	-.0560	.0731	-.7668	.4432
y	<---	x22	-.1506	.1181	-1.2754	.2022
y	<---	x21	-.2931	.1161	-2.5251	.0116
y	<---	x18	-.3030	.1406	-2.1561	.0311
y	<---	x16	.1563	.1019	1.5341	.1250
y	<---	x11	-.0112	.1059	-.1057	.9159
y	<---	x26	-.0330	.0542	-.6077	.5434
y	<---	x27	.1297	.1660	.7809	.4348
y	<---	x28	-.4028	.2368	-1.7010	.0889
y	<---	x29	.0858	.0603	1.4227	.1548
y	<---	x32	-.1894	.1353	-1.3996	.1616
y	<---	x33	.0305	.2113	.1441	.8854
y	<---	x34	.2607	.1158	2.2515	.0244
y	<---	x35	-.0402	.0551	-.7290	.4660
y	<---	x14	.0199	.0743	.2683	.7885
y	<---	x25	.0865	.0681	1.2703	.2040
y	<---	x36	.0498	.0421	1.1843	.2363
y	<---	x2	-.0440	.1046	-.4211	.6737
y	<---	x24	.2111	.1115	1.8931	.0498

We compare the P-value)) with the value of α. If it is less than (α = 0.05), then the variables have a significant effect on y.

Table12: Direct and indirect effects on the internal variable Y

Internal variable	**External variables**	**Direct effects**	**Indirect effects b X_{18}**	**Indirect effects by X_{21}**	**Direct or overall effects**	**Simple Correlations**
Y	X_6	0.2607	0.0324		0.2931	0.232
	X_9	0.2030	0.0581		0.2611	0.226
	X_{18}	-0.1426			-0.1426	-0.202
	X_{21}	-0.1523			-0.1523	-0.229
	X_{24}	0.1130			0.1130	0.107
	X_{34}	0.1354			0.1354	0.074
	X_{14}		0.0247		0.0247	0.034

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

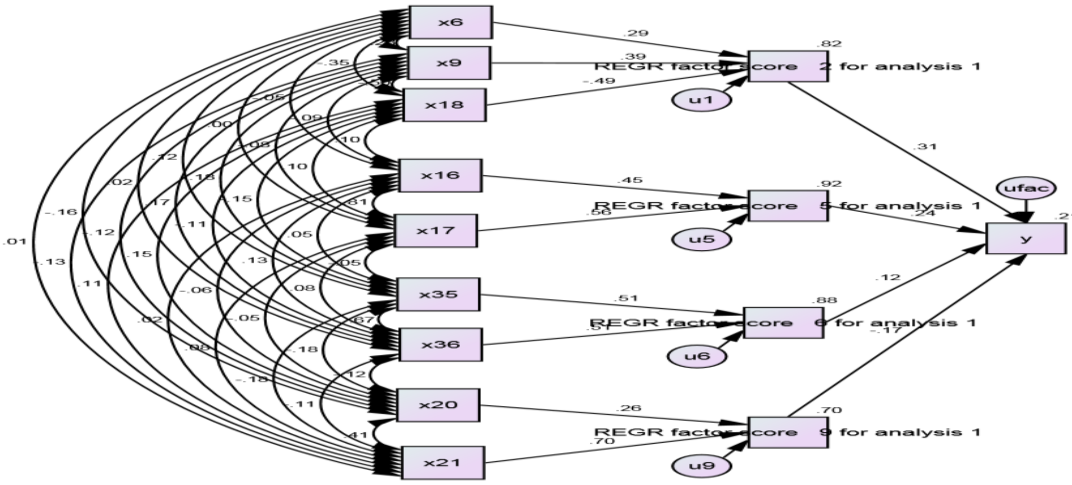
	X_{20}			0.0631	0.0631	-0.234
	X_{37}			0.0182	0.0182	0.017

It is clear from the aforementioned table that the difference between the simple correlation and the total correlation for each variable results from the presence of bilateral correlations between the external variables or the existence of irrational effects. and the value of R^2 for the model is (0.33).

6. The method of harmonization between factor analysis and path analysis

In the beginning, the main compounds are converted into variables, as the factors obtained from the main components are transformed into external variables, and we calculate their effect on the internal variable Y.

After conducting this process, the factors influencing the Y were obtained, after which the path analysis method was applied, then the variables with a non-significant effect were excluded and the significant variables were kept as in the following figure:



The equation represented for Figure (6) is:

$$Y = 0.3102FAC2 + 0.2353FAC5 + 0.1167FAC6 - 0.1655FAC9 + 0.7025ufac \qquad (27)$$

This equation refers to the direct effects of the external variables FAC2, FAC5, FAC6, FAC9 on the internal variable Y, and because the path coefficients show the relative importance, the effect of these variables according to the relative importance is as follows: - FAC2, FAC5, FAC6, FAC9, and this means that the variable FAC2 comes in The first rank has a direct, significant effect of 0.3102, as the poor health condition affects the psychological state of the patient, and thus the blood sugar level rises.

Then followed by the effect of the variable FAC5, which has a direct significant effect of 0.2353, and shows that adherence to the diet leads to a reduction in blood sugar, followed by the effect of the variable FAC6, which has a direct significant effect of 0.1167, as this variable shows that almost half of the study sample do not know the ideal sugar percentage when measuring sugar. In the blood, is it high, low, or ideal in order to take the appropriate action? Finally, the variable FAC9 has a direct, inverse, significant effect of 0.1655. This variable shows that diabetes is not considered a major cause of arterial diseases, but rather a secondary cause. The tests for the significance of the path coefficients are shown in Table 14.

Table 14: Significant tests of the model path coefficients for compatibility

			Estimate	S.E.	C.R.	P	Label
y	<---	FAC2	.2930	.0568	5.1581	***	
y	<---	FAC5	.2223	.0560	3.9705	***	
y	<---	FAC6	.1103	.0569	1.9367	.0489	
y	<---	FAC9	-.1563	.0565	-2.7682	.0056	

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

We note that the value of (P-value) is less than ($\alpha = 0.05$), and therefore all factors have a significant effect on the variable Y.

Table15: Direct effects on the internal variable Y of the model for alignment.

Internal variable	External variables	Direct effects	Indirect effects	Direct or overall effects	Simple Correlations
Y	FAC2	0.3102		0.3102	0.239
	FAC5	0.2353		0.2353	0.235
	FAC6	0.1167		0.1167	0.130
	FAC9	-0.1655		-0.1655	-0.191
	X_6		0.0897	0.0897	0.232
	X_9		0.1199	0.1199	0.226
	X_{18}		-0.1505	-0.1505	-0.202
	X_{16}		0.1055	0.1055	0.234
	X_{17}		0.1309	0.1309	0.225
	X_{35}		0.0597	0.0597	0.134
	X_{36}		0.0600	0.0600	0.152
	X_{20}		-0.0426	-0.0426	-0.234
	X_{21}		-0.1157	-0.1157	-0.229

It is clear from the aforementioned table that the difference between the simple correlation and the total correlation for each variable results from the existence of bilateral correlations between the external variables. The value of R^2 for the proposed model is (0.21).

7. Conclusions

- With regard to what was shown by the results of the factorial analysis by the method of the main components, thirteen factors were obtained that explain 65.368%, as the following variables are strongly related to diabetes because of having the highest saturations, which are [(Did you enter the hospital because of) because most patients are hospitalized due to high blood sugar , (Do you suffer from poor sexual ability) as a large percentage of males in the results of the questionnaire suffer from weak sexual ability due to the fluctuation of the sugar level, (How long do you continue with the diet) as some patients do not adhere to the diet continuously, (Has it changed Your eating habits after having diabetes) since a large proportion of patients do not change their eating habits due to indifference, (what is the ideal sugar ratio before food) since more than half of the study sample does not know the ideal sugar ratio before food, which is 120-140, (what is the ratio The ideal sugar level is two hours after eating) since more than half of the study sample does not know the ideal sugar level before food, which is 160-180, (Do you suffer from an increase in fats) as an increase in fats leads to a rise in blood sugar.
- The following variables are not considered factors associated with diabetes [X_2 (age) as all age groups are susceptible to it, X_8 (is there any family member with diabetes) as the role of heredity in the incidence of diabetes has become less influential than before , X_{10} (Do you see a doctor in order to follow up on your diabetes) as some patients do not have the fees to see a doctor or because of the patient's indifference and negligence in following up on his health condition, X_{13} (What is the most difficult treatment for diabetes that you should follow up on) as the difficulty of the treatment used does not affect Because the patient is required to adhere to the treatment and use it, X_{19} (the type of bread you eat) as barley bread has an effect on blood sugar, but 86% of the study sample used natural bread, X_{22} (do you suffer from poor appetite) as the body needs food to compensate for the deficiency In weight in the case of pill users, but in the case of insulin, the patient needs food in order not to suffer from hypoglycemia, X_{25} (were there any diseases in your eyes during your diabetes) as most of the sample suffers from retinal weakness and may not be due to sugar but rather due to age or because Other diseases.
- With regard to what was shown by the results of the path analysis, six variables were obtained that explain 0.33 of the internal variable Y, and the variables according to importance are [X_6] (how many years have passed since the discovery of your diabetes) as the number of years of infection, as it approaches ten years or more, the probability of developing kidney and eye diseases increases Poor sexual ability, arterial diseases and amputations due to fluctuating blood sugar levels, or due

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023

to the patient's neglect of his health condition and indifference, which causes a permanent rise in sugar, as he suffers from these diseases in an early period of the disease, X_9 (How would you describe your health condition) as most of the sample answered about their health condition It is weak and the explanation for this is that diabetes is closely linked to the patient's psychological state, and therefore the patient's psychological state causes high blood sugar, X_{34} (Do you suffer from amputation) Since diabetes does not result in problems such as amputation of the foot or part of it or one of the fingers except in cases of extreme negligence by the patient and non-compliance with the treatment and diet determined by the attending physician, or due to the doctor's poor follow-up of the disease, X_{24} (Have you had hypoglycemia) Most cases of hypoglycaemia occur due to not eating enough food after taking the treatment or sometimes due to increased physical activity, X_{18} (Do you suffer from numbness and pain in the lower extremities with a reverse effect) as the cause of numbness and pain in the lower extremities is the fluctuation of sugar levels In the blood, it leads to damage to the nerves, X_{21} (Do you suffer from thirst with the opposite effect) Thirst is not a sure indicator of high blood sugar, because it may be due to increased physical activity or high temperatures .

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Bilodeau M, Brenner D. Theory of multivariate statistics. Springer Science & Business Media; 1999.

[2] James GM, Hastie TJ, Sugar CA. Principal component models for sparse functional data. Biometrika 2000;87:587–602.

[3] Sena MM, Almeida MR, Braga JWB, Poppi RJ. Multivariate statistical analysis and chemometrics. Spectroscopic methods in food analysis, CRC Press; 2017, p. 273–314.

[4] Lynn HS, McCulloch CE. Using principal component analysis and correspondence analysis for estimation in latent variable models. Journal of the American Statistical Association 2000;95:561–72.

[5] Kohlmaier JR. A path analysis of binge eating and obesity in African-Americans: Acculturation, racism, emotional distress, binge eating, body dissatisfaction, attitudes towards obesity, dietary restraint, dietary fat intake, and physical activity. Louisiana State University and Agricultural & Mechanical College; 2003.

[6] Branson KM, Agarwal S. Structured Principal Component Analysis. Technical Report, UCSD 2003.

[7] Ota N, Agard DA. Enzyme specificity under dynamic control II: Principal component analysis of α-lytic protease using global and local solvent boundary conditions. Protein Science 2001;10:1403–14.

[8] Calvo RA, Partridge M, Jabri MA. A comparative study of principal component analysis techniques. Proc. Ninth Australian Conf. on Neural Networks, Brisbane, QLD, Citeseer; 1998, p. 276–81.

[9] Rencher AC, Christensen WF. Méthods of multivariate analysis. a john wiley & sons. Inc Publication 2002;727:230–2218.

[10] Patra PSK, Seetharam K, Sahu DP. Development of an expert tool using Principal Component Analysis (PCA) approach: for identifying human diseases & its severity. International Journal of Computer Applications 2010;975:8887.

[11] Doh S-J, Stern RM. Weighted principal component MLLR for speaker adaptation. Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), vol. 88, 1999.

[12] Wiechmann GH, Wiechmann LA. Multiple factor analysis: An approach to attitude validation. The Journal of Experimental Education 1973;41:74–84.

[13] Timm NH. Applied multivariate analysis. Springer; 2002.

[14] Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. Biocomputing 2000, World Scientific; 1999, p. 455–66.

[15] Tarpey T. Self-consistency and principal component analysis. Journal of the American Statistical Association 1999;94:456–67.

Doi: https://doi.org/10.54216/PMTCS.030105
Received: June 22, 2023 Revised: September 15, 2023 Accepted: December 17, 2023