Ridge Regression in Mediation Analysis with an Application

Zakariya Yahya Mohammed Al-Zubaidi   **Assistant. Prof. Dr. Ahmad N. Flaih**

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah

Iraq – 0069 – Wasit   Iraq – 0069 – diwaniyah

zakariyayahya90@yahoo.com   ahmad.flaih@qu.edu.iq

Abstract

This paper came with the aim of knowing the causal mediation analysis and estimation methods in addition to the study data. Mediation is the basis for many fields, where at the present time the use of mediation analysis has become increasingly common in many areas of scientific research as well in other fields such as economic and social as well as medical sciences. Causal mediation analysis is one of the important methods in order to know the mechanism of influence between variables, which has had a large role among researchers in recent times. In order to arrive at a model that leads to accurate estimates, it is necessary to find or search for the method by which the important variables are chosen in order to be included in the model, especially when the study data suffers from the presence of the problem of linearity. Therefore, many methods of estimating the mediation variables have been used, which are the least squares method and Ridge method. In order to apply the estimation methods, a simple random sample of (50) women were drawn to study the factors affecting the number of births (response variable). The study data were analyzed using programming in R language.

Keywords: multi broker form; Indirect effect, direct effect, Least Squares Estimation, ridge regression

انحدار الحرف في تحليل الوساطة مع التطبيق

أ. م. د. أحمد نعيم فليح   زكريا يحيى محمد الزبيدي

قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة القادسية

العراق - 0069 - الديوانية   العراق - 0069 - واسط

ahmad.flaih@qu.edu.iq   zakariyayahya90@yahoo.com

الملخص

جاءت هذه الورقة بهدف معرفة طرق تحليل وتقدير الوساطة السببية بالإضافة إلى بيانـات الدراسـة. الوساطة هـي الأسـاس للعديد مـن المجالات، حيث أصبح استخدام تحليل الوساطة في الوقت الحاضر شائعًا بشكل متزايد في العديد من مجالات البحث العلمي وكذلك فـي مجـالات

أخرى مثل العلوم الاقتصادية والاجتماعية وكذلك العلوم الطبية. يعتبر تحليل الوساطة السببية من الطرق المهمة لمعرفة آلية التأثير بين المتغيرات والتي كان لها دور كبير بين الباحثين في الآونة الأخيرة. من أجل الوصول إلى نموذج يؤدي إلى تقديرات دقيقة، من الضروري إيجاد أو البحث عن الطريقة التي يتم من خلالها اختيار المتغيرات المهمة ليتم تضمينها في النموذج، خاصةً عندما تعاني بيانات الدراسة من وجود مشكلة الخطية. لذلك، تم استخدام العديد من طرق تقدير متغيرات الوساطة، وهي طريقة المربعات الصغرى وطريقة ريدج. لتطبيق طرق التقدير، تم سحب عينة عشوائية بسيطة من (50) امرأة لدراسة العوامل المؤثرة على عدد المواليد (متغير الاستجابة). تم تحليل بيانات الدراسة باستخدام البرمجة بلغة R.

الكلمات الرئيسية: نموذج وسيط متعدد، تأثير غير مباشر، تأثير مباشر، تقدير المربعات الصغرى، انحدار الحرف

1. Introduction

Causal inference is currently an important area of research in statistical sciences there are several different researches using causal Inference, including medicine, economics and sociology causal Inference has become versatile in modern research especially in the medical and health fields (Grotta & Bellocco ,2012).

When there is a relationship between two variables, one independent and the other dependent without any causal explanation explaining the relationship between these two variables in this case comes the role of the researcher in reaching the knowledge of the effect and interpretation of this relationship and this requires the use of the method of mediation analysis model. Therefore, this model is of great importance in the interpretation and knowledge of the causes of the relationship between the cause-and-effect variable depending on the knowledge of one or more mediation variables (Wen, 2013).

Mediation needs to have a third variable called the median let it be (M) and the work of this variable transfer the effect from the independent variable (X) to the dependent variable (Y)

(X ⟶ M ⟶ Y) This is called indirect effect but for the direct effect is the effect of the independent variable (X) direct with dependent variable (Y) (X ⟶ Y) (Murtas, 2016).

Also, when proposing several statistical methods or methods to estimate unknown statistical model parameters if a number of basic assumptions or conditions are met, one of the most famous of these is the least normal squares (OLS) method used to estimate the features of simple and multiple linear regression model due to advantages of this method. It is characterized by the best unbiased linear estimates with the least variation among other unbiased linear estimates. Followed by many researchers who developed different aspects of the problem and methods of solving it, and therefore found ways as possible alternatives to solve this problem gives ease and flexibility in calculations, including the regression of the ridge. So (Hoerl & Kennard, 1970a) added a positive amount between zero and one to the elements of the diameter of the information matrix $(X^T X)$ to solve this problem, they called that constant the Biasing parameter, and the ridge regression method. Ridge regression is a technology

specializing in the analysis of multiple regression data with a multicollinearity problem.

2. Linear Regression

A method known as regression concerned with the relationship between variables. The relationship between variables is formulated by functions.

Suppose that the relationship

$$y = B_0 + B_1 x + \epsilon$$

This equation is the simple linear regression model where y is the dependent variable, x is the predictor variable, and ϵ is the random error. Whereas , B_0 and β are the regression coefficients.

The multiple linear regression model, in concerned with the relationship between the response variable yi and many predictors, $xi = (x_{i1}, \dots \dots x_{ik})$, for a given n samples, $(x_i, y_i)_{i=1}^n$.

The multiple model is

$$yi = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i, \qquad i = 1,2, \dots, n. \quad (1)$$

The coefficients, $\beta_0, \beta_1, \dots, \beta_k$, are unknown, and $\epsilon_1, \dots, \epsilon_n$, are distributed according to $N(0, \sigma^2)$.

In a matrix form. the model (2.1) can be written as follows:

$$Y = X\beta + \epsilon, \qquad\qquad (2)$$

Where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \beta$$

$$= \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

So, we have

$$E(\epsilon) = 0 \quad and \quad Cov(\epsilon) = \sigma^2 I_n, \qquad (3)$$

Least squares estimation is the well know method to estimates the unknow vector of B.

3. Least Squares Estimation

The $\hat{\beta}$ in LSE is defined by (Montgomery et al. ,2013).

$$\hat{\beta} = \underset{\beta}{argmin} \ RSS(\beta)$$

Where RSS is the residuals sum of squares , i.e.,

$$RSS(\beta) = \sum_{i=1}^n (yi - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}))^2 \qquad (4)$$

So, the LSE $(\hat{B})$ is

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

4. Collinearity

when the **X** matrix have full rank and the $(X^T X)$ matrix is nearly singular. The multi-collinearity is appear because the columns of **X** correlated so that they are linear dependent. (Gentle, 2007). The effect of multi-collinearity is to make the variances of the parameter estimates inflate.

5. Ridge Regression

In recent times, shrinkage techniques are highly accurate and parameter selection is of great importance. The settlement of Tikhonov, a second name for RR, deals with many predictors and a matrix of an unconditional form of X (since $X^T X$ is not reversible or close to singular). Whereas, installing a model with multiple forecasts with no adjustment results in an LSE solution. In addition, LSE relies on $(X^T X)^{-1}$ i.e., if the order $(X) < k$, then $(X^T X)$ has no inverse. However, RR has the potential to overcome these barriers by restricting the coefficient estimates; thus, it Can reduce resident variance and provide some bias, (James, 2013).

RR coefficients are estimated by minimization, similar to LSE:

$$S(\boldsymbol{\beta}, \lambda) = \mathbf{RSS}(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{k} \beta_j^2 \qquad (5)$$

over β for a given λ, where is called λ a tuning parameter and $\lambda \geq 0$. Note the $RSS(\beta)$ defined in (4). The (RRE) ridge regression estimator is denoted by $\hat{\beta}^R(\lambda)$.

Tikhonov works to find acceptable estimates of the data, by making RSS (β) small, as well as the term, $\lambda \sum_{j=1}^{k} \beta_j^2$, it also becomes small when shrinking estimates $(\hat{\beta}_1, \ldots, \hat{\beta}_k)$ are approximately zero (James et al., 2013). We define (5) in the representation of the vector and vector norm, and their shape will be as follows:

$$\min_{\beta \in R^k} \quad \|Y - X\beta\|^2 \qquad (6)$$

$$subject\ to \qquad \|\beta\|_2^2 \leq r^2$$

$$\|\beta\|_2^2 = \sum_{j=1}^{k} \beta_j^2 \ and \ r^2 > 0,$$

$and\ r^2$ be related to λ.

The setting parameter λ plays a very important role here. Controls the relative influence of the two terms in $S(\beta, \lambda)$. For example, when the value of $\lambda = 0$, it becomes $RR\ LSE$, but when λ gets bigger, $\|\hat{\beta}^R\|_2^2$ shrinks; thus, the zero coefficient estimates approach RR. And thus, different from LSE, for each value λ, we obtain a different set of parameter estimates, $\hat{\beta}^R(\lambda)$.

When looking at $S(\beta, \lambda)$, we notice that the contraction penalty, $\lambda \sum_{j=1}^{k} \beta_j^2$, is applied only to $\beta_1, \beta_2, \ldots, \beta_k$ and does not apply to β_0. The term intersection is the measurement of the average value of the response if $x_{i1} = x_{i2} = \cdots = x_{ik} = 0$. To delete β_0, we standardize the scales, which means that each column is in the middle so that $\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0$ and $\frac{1}{n}\sum_{i=1}^{k} x_{ij}^2 = 1$, , for $j = 1, 2, \ldots, k$. Here we note that standardization occurs if the features in units differ. For convenience, we also focus on outcome values like $\frac{1}{n}\sum_{i=1}^{k} y_i = 0$. Once we get the RR coefficient estimates, we can retrieve the term intercept by

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{k} \bar{x}_j \hat{\beta}_j^R(\lambda)$$

Where that $\bar{y}$ and $\bar{x}_j$, because $j = 1, 2, \ldots, k$, is the original method. Using notation vector and vector norm, we can write an RR problem like

$$\min_{\beta \in R^k} S(\beta, \lambda) = \min_{\beta \in R^k} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \quad for\ som \quad \lambda \geq 0,$$

Where $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ and $\|\beta\|_2^2 = \sum_{j=1}^{k} \beta_j^2$. Here the matrix X does not include a column of them. This equation (5) is referred to as Lagrangian from RR.

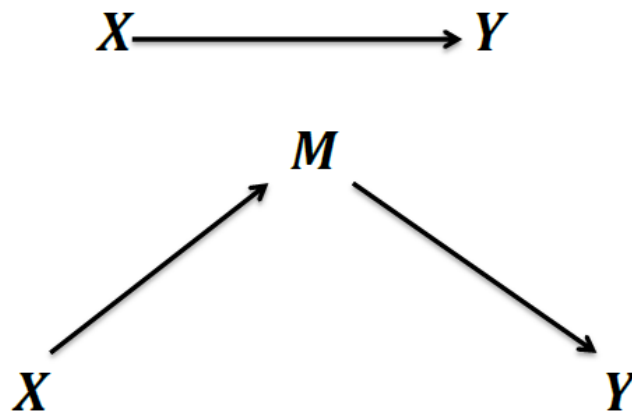We can give another format for the RR problem as follows:

$$S(\beta, \lambda) = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{k} x_{ij}\beta_j)^2 + \sum_{j=1}^{k}(0 - \sqrt{\lambda\beta_j})^2 \quad (7)$$

This method reconfigures another LSE problem for a set of extended data by adding k views, which results in an expanded sample matrix X_λ and Y_λ below:

$$X_\lambda = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \\ \sqrt{\lambda} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda} \end{bmatrix}, \quad Y_\lambda = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

It is clear

$$X_\lambda = \begin{bmatrix} X \\ \sqrt{\lambda}I_k \end{bmatrix}, \quad Y_\lambda = \begin{bmatrix} Y \\ 0 \end{bmatrix}$$

Which ultimately results in the solution of RR, (Hoerl and Kennard ,1970),

$$\hat{\beta}^R(\lambda) = (X_\lambda^T X_\lambda)^{-1} X_\lambda^T Y_\lambda = (X^T X + \lambda I_k)^{-1} X^T Y, \quad (8)$$

In the form of LSE.

6. Mediation analysis

Most of the research used in many fields focuses on the relationship between two variables, where the independent variable X (cause variable) and the other variable is deliberate variable Y (response) when we add a third variable (M) may be difficult then a new variable called median and be Between the independent variable (X) and the dependent variable (Y) called the mediation variable (M) , which explains the effect of the cause variable X on the result variable Y, then the independent variable (X) is the cause of the mediation variable (M), which causes the result variable (Y) $(X \rightarrow M \rightarrow Y)$ (D.P.Macimon & Fairchild & Fritz, 2007) (Grotta & Bllocco, 2012).

Figure (1) Shows the effects of the variables before and after adding the mediation variable (M).

7. Single - Mediator Model

The individual mediation model is one of the simplest models in the analysis of causal mediation. This model consists of three variables X is the independent variable (treatment variable), M mediation variable, Y dependent variable (result variable) so this model is used when the independent variable is changed to This one intermediate variable leads to an effect in the result. This model (Individual Mediator Model), also called individual mediation or simple mediation, has been applied in several areas, including behavioral, social, and educational sciences, which makes the researcher interested to learn about the side effects not studied (wen, 2013).

7.1 Regression equations used to evaluate mediation

Baron & Kenny (1986) is one of the most common methods used in the analysis of mediation. This was based on the modeling of linear structural equation (LSEM) in order to provide a mediation approach to an individual median model where the following regression equations are required:

$$Y = r_1 + cX + e_1 \quad \ldots\ldots\ldots\ldots\ldots. (9)$$

$$M = r_2 + aX + e_3 \quad \ldots\ldots\ldots\ldots. (10)$$

$$Y = r_3 + c'X + bM + e_2 \quad \ldots\ldots. (11)$$

Whereas:

Y: represents the result of interest.

X: represents the independent variable (treatment variable).

M: Represents the mediation variable.

C: represents the relationship between the processing variable and the result in equation (9) where the total effect is called.

C': The parameter that binds as the process variable with the result variable under the influence of the mediation variable is represented in equation (11) where the partial effect is called.

a: The parameter that transmits the effect of the processing variable on the mediation variable is in equation (10).

b: The parameter that transmits the effect of the mediation variable on the result variable in equation (11).

r_1, r_2, r_3: objections to each equation.

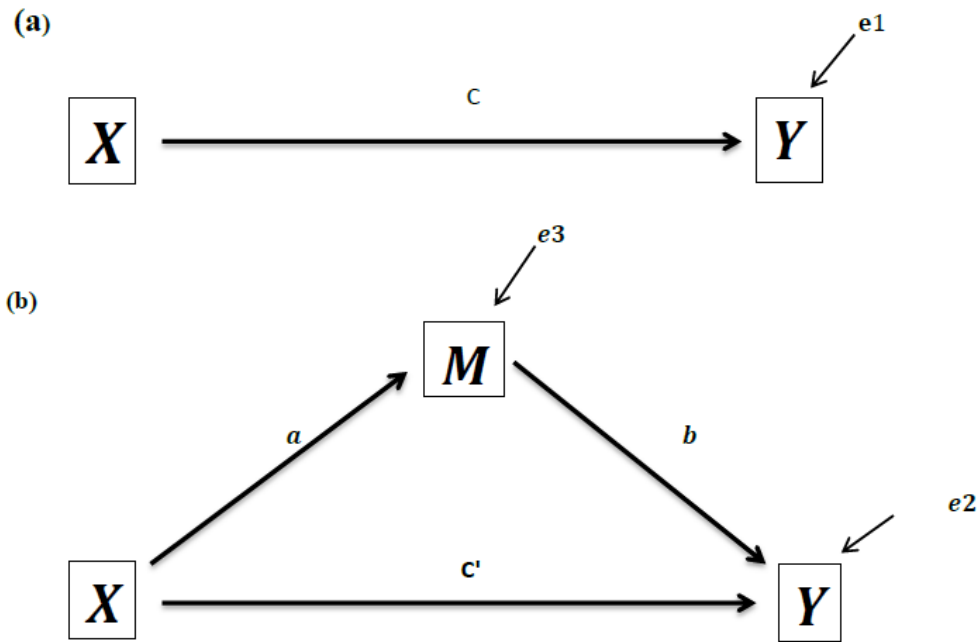e_1, e_2, e_3: errors.

$Equation\ (9)$: represents the total effect in Figure (A).

$Equations\ (10), (11)$: Determine the mediation model in Figure (B).

(D.P Mackinnon & pilot , 2015) (Pachali & Mayer & landwehr , 2018)

(Tofighi & Thoemmes , 2014) .

Figure (2)   **(a) -** The direct effects of the X variable on the Y variable .

(b) - Indirect effects of the variable X on the variable Y by the mediation variable M .

The symbols (a, b, c) are considered of particular importance, where (a) denotes the effect of the treatment variable (X) on the mediation variable (M) , (b) denotes the effect of the mediation variable (M) on the result variable (Y) , (c) denotes the effect of the treatment variable (X) on the outcome variable (Y) (David Mackinnon, 2012) (D.P.Mackinnon, 2007) (D.P.Mackinnon & Pirlott, 2015).

8. Multiple Mediator Model

After we describe the individual mediation model, we turn in this part to a broader and more complex mediation model than the previous model, which is a multi-mediation model with a description of its equations with its graphical forms.

8.1 Regression equations used to evaluate mediation

In order to provide a mediation approach to the multi-median model, the following regression equation is required:

$$Y = r_1 + cX + e_1 \qquad …………. (12)$$

$$Y = r_2 + c'X + b_1M_1 + … … + b_nM_n + e_2 \ .(13)$$

$$M_1 = r_3 + a_1X + e_3 \qquad …………… (14)$$

$$M_n = r_n + a_nX + e_n \qquad …………. (15)$$

whereas

Y: represents the result variable.

X: represents a treatment variable.

$M_1, \ldots, M_n$: Causal Mediation Variables.

C: The parameter that binds the treatment variable with the outcome variable represents the absence of mediation variables.

C': The parameter that binds the treatment variable to the outcome variable is the effect of the mediation variables.

a_1: The parameter that associates the transaction variable with the first broker variable.

a_n: The parameter that links the transaction variable to other mediation variables.

b_1: The parameter that binds the first mediation variable to the result variable

b_n: The parameter that links the other mediation variables to the result variable.

$r_1, r_2, r_3, \ldots, r_n$: are objections in each equation

$e_1, e_2, e_3, \ldots, e_n$: are errors.

Equation (12) is the estimation of the direct effect in the absence of mediation.

Equations (13)(14)(15) is the estimation of the indirect effect by the presence of the mediation model (David Mackinnon, 2012).
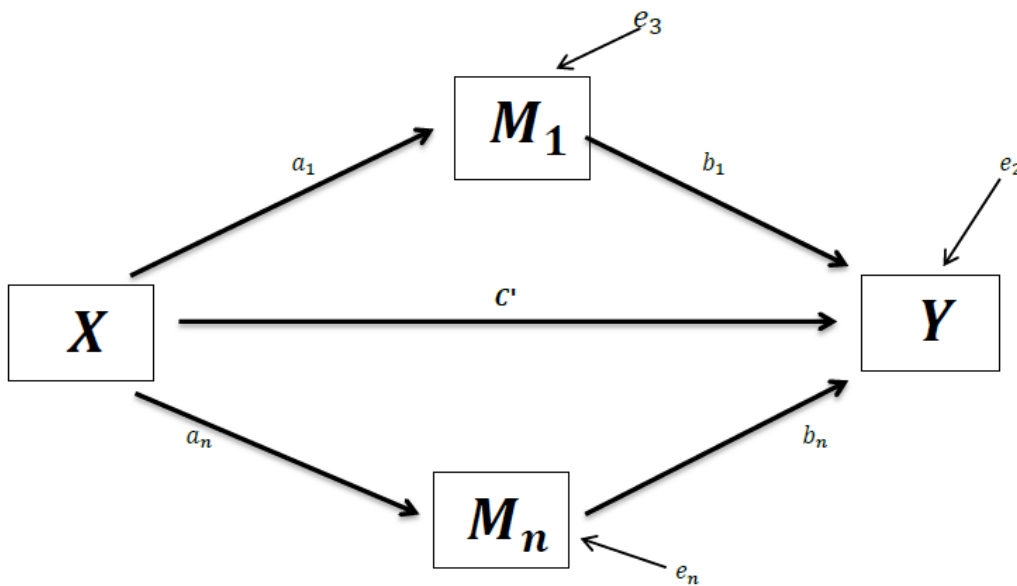
Figure (3) : Represents a diagram of the multi-mediation model using two variables, so the independent variable (X) is related to two arguments (M_1) and (M_n) which are related to the dependent variable (Y). Note that there are several other possible mediation models, for example X to M_1 to M_n to y.

9. The Total Effect

The multiple mediation model in Figure (3) consists of two mediation variables whose total effect is the direct effect represented by parameter (C') and the indirect effect of the mediation variables $(M_1, \dots, M_n)$ (Hayes, 2009).

$$Total\ effect = C' + a_1 b_1 + \dots + a_n b_n\ ..(16)$$

10. Effect of mediation

We divide the total effect into direct effects, which are represented by the parameter (C') and others that are indirect and that are represented by the effects of the mediator $(a_1 b_1, \dots, a_n b_n)$ in the analysis of mediation, so it should be:

$$a_1 b_1 + \dots + a_n b_n = C + C' \qquad (17)$$

In the case of an unequal reason, this is due to the difference in the sample sizes between the equations Therefore, the parameters of the above models are estimated using different methods, such as (OLS) (Preacher & Hayes, 2008).

11. Testing the importance and confidence intervals of mediation effects

One of the most important methods used to test the significance of the mediation effect (ab) is to estimate the standard error and compare the resulting Z degrees with the critical value of the standard normal distribution (D.P.Mackinnon et al. ,2004).

The standard error and estimated causal mediation effect can also be used to establish confidence intervals for mediation effect. It is well known that confidence intervals use standard error in estimation, which is why we believe that the confidence intervals used may provide a number of effect values rather than a single value. Therefore, confidence intervals are tools with common uses in research because they ask the researcher to look for the value of the effect as well as its statistical significance (Harlow,nd) and to test the importance of the indirect effect (mediator), So we need to find the standard error of the sample by the median (ab) The most commonly used tests to estimate the standard errors of indirect effect are:

11.1 Strategy for causal action to establish mediation for the multi-median model

The multi-mediation form contains a set of steps:

1- The treatment variable (X) affects the outcome variable (Y) through (C) as in equation (12).

2- The treatment variable (X) affects the mediation of $(M_1, \dots, M_n)$ by (a_1) and (a_n) as in equations (14)(15).

3- Mediation variables affect the outcome variables (Y) after controlling the treatment variable (X) factors $(b_1, \dots, b_n)$ as in equation (13).

4- The treatment variable must have an unimportant effect (direct effect C') on the outcome variable in order to achieve full mediation as found in equation (5), but if there is an effect of the treatment variable on the outcome variable (C') here is a partial mediation (Wen , 2013).

11.2 Product of parameter coefficients testing

The standard error for the internal effect of a multiple mediation model is as follows (David Mackinnon , 2012).

$$S_{a1\,b1} = \sqrt{S_{a1}^2\,b_1^2 + S_{b1}^2\,a_1^2} \qquad (18)$$

Another variation of the standard error of the multiple mediation model can also be used:

$$S_{a1\,b1} + \cdots + S_{an\,bn} =$$
$$\sqrt{\begin{array}{c} S_{a1}^2 S_{b1}^2 + S_{b1}^2 a_1^2 + \cdots + \\ S_{an}^2 b_n^2 + S_{bn}^2 a_n^2 + 2a_1 \ldots a_n\ S_{b1 \ldots bn} \end{array}} \qquad (19)$$

Rewrite the equation above as follows:

$$S_{a1\,b1} + \cdots + S_{an\,bn} = \sqrt{\begin{array}{c} S_{a1\,b1}^2 + \cdots + S_{an\,bn}^2 \\ +2a_1, \ldots, a_n\ S_{b1,\ldots,bn} \end{array}} \qquad (20)$$

$(2b_1, \ldots, b_n\ S_{a1,\ldots,an})$ Must be added to the above equations when there is a non-zero difference between $(a_1, \ldots, a_n)$

$(S_{b1,\ldots,bn})$: Represents the common difference.

When the overall effect of the medium is present, the standard error is as follows:

12. Estimation Methods

12. 1 The Method Of Least Squares

This topic will include estimating the expected average of the response variable using the available information from the predictors variables $(X_1, X_2, \ldots, X_{19})$ through the use of the three regression models, namely:

First Model: Estimating the linear relationship of the regression   model (Y/ X)

Here the parameters of the predicted variables (X) will be estimated, and the expected mean of the response variable (Y) will be estimated through regression model (12).

Where it will be written as follows:

$$yi = r_1 + \sum_{i=1}^{19} C_i X_i + e_1 \qquad (21)$$

where the vector (C) represents the total or direct effect of the predictors variables (X) on the response variable (Y) and the hypothesis test $H_0: \underline{C} = 0$,   and the table below (1) shows a summary of statistics for the estimated model.

Table (1) shows a summary of the estimated model statistics

Variable	Estimate	S.E.	t	P-Value
X_1	1.875	2.363e-01	7.936	7.39e-09 ***
X_2	-1.562	1.864e-01	-8.379	2.37e-09 ***
X_3	2.072e-01	1.289e-01	1.607	0.11858
X_4	-1.785e-01	1.394e-01	-1.281	0.21008
X_5	-9.483e-02	1.673e-01	-0.567	0.57502
X_6	1.013e-01	1.302e-01	0.778	0.44273

X_7	-1.974e-01	1.275e-01	-1.548	0.13200
X_8	2.154e-01	1.414e-01	1.523	0.13814
X_9	1.841e-01	1.128e-01	1.632	0.11306
X_{10}	2.941e-01	1.017e-01	2.893	0.00705 **
X_{11}	-1.094e-01	1.020e-01	-1.073	0.29198
X_{12}	2.200e-01	1.058e-01	2.078	0.04634 *
X_{13}	1.253e-01	1.044e-01	1.200	0.23967
X_{14}	-3.432e-02	1.259e-01	-0.273	0.78699
X_{15}	8.100e-03	1.150e-01	0.070	0.94432
X_{16}	-8.956e-02	1.187e-01	-0.754	0.45659
X_{17}	6.112e-02	1.499e-01	0.408	0.68645
X_{18}	-5.024e-02	9.762e-02	-0.515	0.61055
X_{19}	-1.551e-01	1.233e-01	-1.258	0.21816

Table (1) shows the estimated values of the model parameters (12) in addition to the standard errors and the test (t) and the values of (P-Value) where we notice that the variables (Age of the woman) (Woman's age at marriage) (Number of deceased children) (Woman with thyroid disease) have a significant effect level of significance ($\alpha = 0.05$) and this is clear from the values of the column (P-Value) as well as we note from this The table shows that the rest of the variables do not have a significant effect in the study model, and this is evident from the values of (P-Value) in comparison with the level of significance ($\alpha = 0.05$).

To test the significance of the estimated model (12) by using the least squares method, we will test the hypothesis.

$H_0: Ci = 0$ and $H_1: Ci \neq 0$ $i = 1, ... , 19$

Where the table below (2) shows the analysis of variance table for the estimated model.

Table (2) analysis of variance

S.O.V	D.F	Sum Sq	Mean Sq	F value	Pr(>F)	R^2	Adjusted R^2
Model	19	76.424	4.0223	11.999	2.916e-09***	0.8837	0.8101
Residuals	30	10.056	0.3352				
Total	49	86.480					

From the table (2) We note from the value of (Pr(>F)) the significance of the model estimated using the method of least squares, as its value was close to zero and it is less than the level of significance ($\alpha = 0.05$) and thus rejected H_0.

We also note that the value of (R^2) is equal to (%88), which is interpreted as the ratio of the total change occurring in the response variable (number of children born) as a result of its effect on the values of the predictors variables (X), where the value of (R^2) can be considered as an indicator of the suitability of the estimated model or an indicator of the strength of the linear relationship between the number of children born and predictors variables.

We also note the value of ($Adjusted\ R^2$), which is equal to (%81), which is another measure to judge the extent to which the estimated model

matches the data, assuming a change in the number of predictors variables. We can use ($Adjusted\ R^2$) to compare models that have different numbers of predictors variables.

Second Model: Estimating the Regression Model $(Y/X, M)$

In this model (13) the parameters of the predictors variables (X) will be estimated ($C' =$ partial effect) and the expected estimate of the response variable (Y) with another direct estimate called (the mediation effect) by the parameter (b) which links the mediation variable to the outcome variable in order to test the partial effect (C') , We will test the following null hypothesis : $H_0: \underline{C'} = 0\ and\ H_0: b = 0$, and the table (3) below shows a summary of statistics for the estimated model.

Table (3) shows a summary of the estimated model statistics

Variable	Estimate	S.E.	t	P-Value
X_1	1.898	2.343e-01	8.102	1 6.20e-09 ***
X_2	-1.667	2.015e-01	-8.275	4.01e-09 ***
X_3	2.279e-01	1.285e-01	1.773	0.0867
X_4	-1.859e-01	1.380e-01	-1.347	0.1883
X_5	-9.658e-02	1.654e-01	-0.584	0.5638
X_6	1.236e-01	1.299e-01	0.951	0.3493
X_7	-1.958e-01	1.261e-01	-1.553	0.1312
X_8	2.511e-01	1.425e-01	1.762	0.0886
X_9	2.299e-01	1.170e-01	1.965	0.0590
X_{10}	1.603e-01	1.440e-01	1.114	0.2746
X_{11}	-1.249e-01	1.016e-01	-1.230	0.2286

X_{12}	2.175e-01	1.047e-01	2.077	0.0467 *
X_{13}	1.410e-01	1.040e-01	1.356	0.1855
X_{14}	-7.211e-02	1.278e-01	-0.564	0.5770
X_{15}	3.036e-02	1.150e-01	0.264	0.7937
X_{16}	-1.037e-01	1.179e-01	-0.879	0.3865
X_{17}	4.531e-02	1.488e-01	0.305	0.7629
X_{18}	-1.043e-01	1.051e-01	-0.992	0.3293
X_{19}	-3.048e-01	1.678e-01	-1.816	0.0797
M	2.439e-01	1.880e-01	1.298	0.2046

Table (3) shows the estimated values of the model parameters (13) in addition to the standard errors and the test (t) and the values of (P-Value) where we notice that the variables (Age of the woman) (Woman's age at marriage) (Woman with thyroid disease) have a significant effect level of significance ($\alpha = 0.05$) and this is clear from the values of the column (P-Value) as well as we note from this The table shows that the rest of the variables do not have a significant effect in the study model, and this is evident from the values of (P-Value) in comparison with the level of significance ($\alpha = 0.05$).

To test the significance of the estimated model (13) by using the least squares method, we will test the hypothesis.

$H_0: Ci = 0$ and $H_1: Ci \neq 0$ $i = 1, ... ,19$

$H_0: b = 0$ and $H_1: b \neq 0$

Where the table below (4) shows the analysis of variance table for the estimated model.

Table (4) analysis of variance

S.O.V	D.F	Sum Sq	Mean Sq	F value	Pr(>F)	R^2	Adjusted R^2
Model	20	76.976	3.8488	11.744	4.865e-09 ***	0.8901	0.8143
Residuals	29	9.504	0.3277				
Total	49	86.480					

From the table (4) We note from the value of (Pr(>F)) the significance of the model estimated using the method of least squares, as its value was close to zero and it is less than the level of significance ($\alpha = 0.05$) and thus rejected H_0.

We also note that the value of (R^2) is equal to (%89), which is interpreted as the ratio of the total change occurring in the response variable (number of children born) as a result of its effect on the values of the predictors variables (X) by the parameter (b) which links the mediation variable to the outcome variable in order to test the partial

effect (C'), where the value of (R^2) can be considered as an indicator of the suitability of the estimated model or an indicator of the strength of the linear relationship between the number of children born and predictors variables.

We also note the value of ($Adjusted\ R^2$), which is equal to (%81), which is another measure to judge the extent to which the estimated model matches the data, assuming a change in the number of predictors variables. We can use ($Adjusted\ R^2$) to compare models that have different numbers of predictors variables.

Third Model: Estimating the Regression Model (M/X)

In this model (14), the effect of the parameters of the predictor's variables (X) on the mediation variable (M) will be estimated in order to test the direct effect (a).

Null hypothesis: $H_0: a = 0$

and the table (5) below shows a summary of statistics for the estimated model.

Table (5) shows a summary of the estimated model statistics

Variable	Estimate	S.E.	t	P-Value
X_1	-6.357e-02	1.527e-01	-0.416	0.6801
X_2	2.917e-01	1.204e-01	2.422	0.0217 *
X_3	-5.705e-02	8.331e-02	-0.685	0.4987
X_4	2.022e-02	9.007e-02	0.224	0.8239
X_5	4.825e-03	1.081e-01	0.045	0.9647
X_6	-6.144e-02	8.413e-02	-0.730	0.4709
X_7	-4.346e-03	8.238e-02	-0.053	0.9583
X_8	-9.840e-02	9.137e-02	-1.077	0.2901
X_9	-1.262e-01	7.288e-02	-1.732	0.0936
X_{10}	3.689e-01	6.569e-02	5.616	4.10e-06 ***
X_{11}	4.277e-02	6.591e-02	0.649	0.5214
X_{12}	6.950e-03	6.840e-02	0.102	0.9197
X_{13}	-4.335e-02	6.749e-02	-0.642	0.5256
X_{14}	1.042e-01	8.133e-02	1.282	0.2098
X_{15}	-6.140e-02	7.432e-02	-0.826	0.4152
X_{16}	3.894e-02	7.673e-02	0.507	0.6155
X_{17}	4.362e-02	9.689e-02	0.450	0.6558
X_{18}	1.492e-01	6.308e-02	2.365	0.0247 *
X_{19}	4.128e-01	7.966e-02	5.182	1.39e-05 ***

Table (5) shows the estimated values of the model parameters (14) in addition to the standard errors and the test (t) and the values of (P-Value) where we notice that the variables (Woman's age at marriage) (Number of deceased children) (Match the blood) (Gestational diabetes) have a significant effect level of significance ($\alpha = 0.05$) and this is clear from the values of the column (P-Value) as well as we note from this The table shows that the rest of the variables do not have a significant effect in the study model, and this is evident from the values of (P-Value) in comparison with the level of significance ($\alpha = 0.05$).

To test the significance of the estimated model (14) by using the least squares method, we will test the hypothesis.

$H_0: a = 0$ and $H_1: a \neq 0$ $i = 1, \dots, 19$

Where the table below (6) shows the analysis of variance table for the estimated model.

Table (6) analysis of variance

S.O.V	D.F	Sum Sq	Mean Sq	F value	Pr(>F)	R^2	Adjusted R^2
Model	19	17.9813	0.94639				
				6.762	2.218e-06 ***	0.8107	0.6908
Residuals	30	4.1987	0.13996				
Total	49	22.1800					

From the table (6) We note from the value of (Pr(>F)) the significance of the model estimated using the method of least squares, as its value was close to zero and it is less than the level of significance ($\alpha = 0.05$) and thus rejected H_0.

We also note that the value of (R^2) is equal to (%81), which is interpreted the effect of the parameters of the predictors variables (X) on the mediation variable (M) which will be estimated in order to test the direct effect (a), where the value of (R^2) can be considered as an indicator of the suitability of the estimated model or an indicator of the strength of the linear relationship between the number of children born and predictors variables.

We also note the value of ($Adjusted\ R^2$), which is equal to (%69), which is another measure to judge the extent to which the estimated model matches the data, assuming a change in the number of predictors variables. We can use ($Adjusted\ R^2$) to compare models that have different numbers of predictors variables.

12.2 Ridge Regression Method

Ridge regression provides us with an alternative style or method for estimating, which means that it is an alternative estimation method

for model parameters when we find that the predictors variables have high correlations between them. Depending on the correlations matrix, we find that there are variables that have high correlations between them, which means that there is a problem of linear multiplicity, as the least squares method is known to be ineffective in the presence of the problem of linear multiplicity, so we will resort to the Ridge method in estimating the parameters.

First Model: Estimating the linear relationship of the regression model (Y/ X)

In this model (21) the parameters of the predictors variables (X) will be estimated and the expected mean of the response variable (Y) will be estimated through regression model (21) , as the vector (C) represents the total or direct effect of the predictors variables (X) on the response variable (Y) and the hypothesis test $H_0: \underline{C} = 0$, and the table (7) below shows a summary of statistics for the estimated model.

Table (7) shows a summary of the estimated model statistics

Variable	Estimate	S.E.	t	P-Value
X_1	1.803	3.301e-01	5.461	4.73e-08 ***
X_2	-1.033	3.823e-01	-2.702	0.00690 **
X_3	-4.344e-02	4.467e-01	-0.097	0.92253
X_4	-1.159e-01	4.272e-01	-0.271	0.78614
X_5	1.343e-01	3.942e-01	0.341	0.73336
X_6	-3.874e-01	4.435e-01	-0.874	0.38238
X_7	-6.236e-01	4.435e-01	-1.406	0.15973
X_8	1.176	4.154e-01	2.829	0.00466 **
X_9	8.304e-02	4.625e-01	0.180	0.85750
X_{10}	1.890	4.726e-01	3.999	6.35e-05 ***
X_{11}	5.599e-01	4.686e-01	1.195	0.23214
X_{12}	5.370e-01	4.597e-01	1.168	0.24272
X_{13}	3.367e-01	4.668e-01	0.721	0.47077
X_{14}	6.738e-01	4.460e-01	1.511	0.13088
X_{15}	5.078e-02	4.594e-01	0.111	0.91197
X_{16}	5.939e-01	4.537e-01	1.309	0.19054
X_{17}	-2.063e-01	4.119e-01	-0.501	0.61647
X_{18}	-5.759e-01	4.762e-01	-1.209	0.22656
X_{19}	6.277e-01	4.569e-01	1.374	0.16953

Table (7) shows the estimated values of the model parameters (21) in addition to the standard errors and the test (t) and the values of (P-Value) where we notice that the variables (Age of the woman) (Woman's age at marriage) (The age of the husband) (Number of deceased children) have a significant effect level of significance ($\alpha = 0.05$)

and this is clear from the values of the column (P-Value) as well as we note from this The table shows that the rest of the variables do not have a significant effect in the study model, and this is evident from the values of (P-Value) in comparison with the level of significance ($\alpha = 0.05$).

To test the significance of the estimated model (21) by using the ridge method, we will test the hypothesis $H_0: Ci = 0$ and $H_1: Ci \neq 0$
$i = 1, \dots, 19$

Where the table below (8) shows the analysis of variance table for the estimated model.
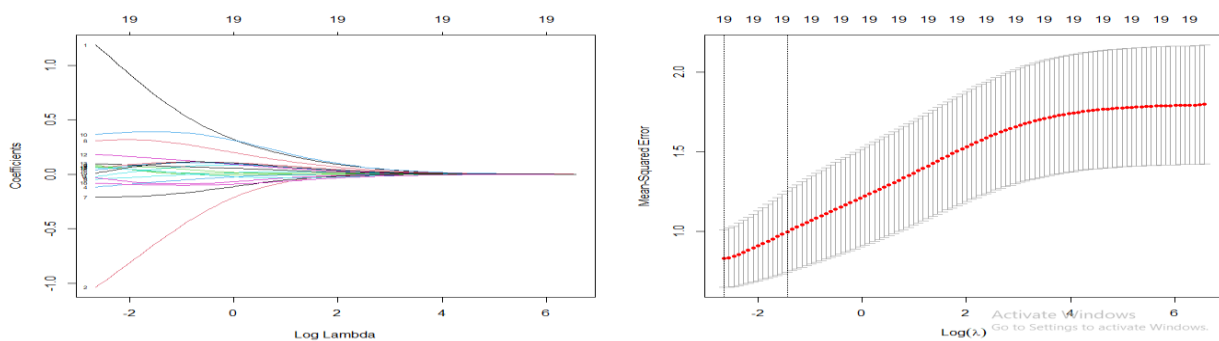
Table (8) analysis of variance

S.O.V	D.F	Sum Sq	Mean Sq	F value	Pr(>F)	R^2	Adjusted R^2
Model	19	75.23	3.95947	10.5586	1.385924e-08	0.8699	0.7875
Residuals	30	11.25	0.375				
Total	49	86.48					

From the table (8) We note from the value of (Pr(>F)) the significance of the model estimated using the method of ridge, as its value was close to zero and it is less than the level of significance ($\alpha = 0.05$) and thus rejected H_0.

We also note that the value of (R^2) is equal to (%86), which is interpreted as the ratio of the total change occurring in the response variable (number of children born) as a result of its effect on the values of the predictors variables (X), where the value of (R^2) can be considered as an indicator of

the suitability of the estimated model or an indicator of the strength of the linear relationship between the number of children born and predictors variables.

We also note the value of ($Adjusted\ R^2$), which is equal to (%78), which is another measure to judge the extent to which the estimated model matches the data, assuming a change in the number of predictors variables. We can use ($Adjusted\ R^2$) to compare models that have different numbers of predictors variables.

Figure (4) a schematic diagram of the first model using the ridge method

Second Model: Estimating the Regression Model $(Y/X, M)$

In this model (13) the parameters of the predicted variables (X) will be estimated and the

expected estimate of the response variable (Y) with another direct estimate called (the mediation effect) by the parameter (b) which links the mediation variable to the outcome variable in order

to test the partial effect (C') , null hypothesis : $H_0 : \underline{C'} = 0 \ and \ H_0 : b = 0$ and the table (9) below shows a summary of statistics for the estimated model.

Table (9) shows a summary of the estimated model statistics

Variable	Estimate	S.E.	t	P-Value
X_1	1.576	3.024e-01	5.211	1.88e-07 ***
X_2	-9.304e-01	3.494e-01	-2.663	0.00774 **
X_3	-3.064e-02	4.162e-01	-0.074	0.94132
X_4	-7.819e-02	3.977e-01	-0.197	0.84415
X_5	1.160e-01	3.670e-01	0.316	0.75202
X_6	-3.143e-01	4.116e-01	-0.763	0.44518
X_7	-5.773e-01	4.131e-01	-1.397	0.16229
X_8	1.088	3.841e-01	2.833	0.00461 **
X_9	9.107e-02	4.315e-01	0.211	0.83286
X_{10}	1.637	4.165e-01	3.930	8.48e-05 ***
X_{11}	5.365e-01	4.365e-01	1.229	0.21903
X_{12}	4.710e-01	4.269e-01	1.103	0.26992
X_{13}	3.137e-01	4.360e-01	0.720	0.47181
X_{14}	6.153e-01	4.140e-01	1.486	0.13725
X_{15}	6.966e-02	4.265e-01	0.163	0.87025
X_{16}	5.856e-01	4.232e-01	1.384	0.16645
X_{17}	-2.205e-01	3.820e-01	-0.577	0.56380
X_{18}	-5.766e-01	4.425e-01	-1.303	0.19261
X_{19}	4.420e-01	3.942e-01	1.121	0.26216
M	6.310e-01	3.621e-01	1.743	0.08138

Table (9) shows the estimated values of the model parameters (13) in addition to the standard errors and the test (t) and the values of (P-Value) where we notice that the variables (Age of the woman) (Woman's age at marriage) (The age of the husband) (Number of deceased children) have a significant effect level of significance ($\alpha = 0.05$) and this is clear from the values of the column (P-Value) as well as we note from this The table shows that the rest of the variables do not have a significant effect in the study model, and this is evident from the values of (P-Value) in comparison with the level of significance ($\alpha = 0.05$).

To test the significance of the estimated model (13) by using the ridge method, we will test the hypothesis.

$$H_0: Ci = 0 \quad and \quad H_1: Ci \neq 0 \qquad i = 1, \ldots, 19$$

$$H_0: b = 0 \quad and \quad H_1: b \neq 0$$

Where the table below (10) shows the analysis of variance table for the estimated model.
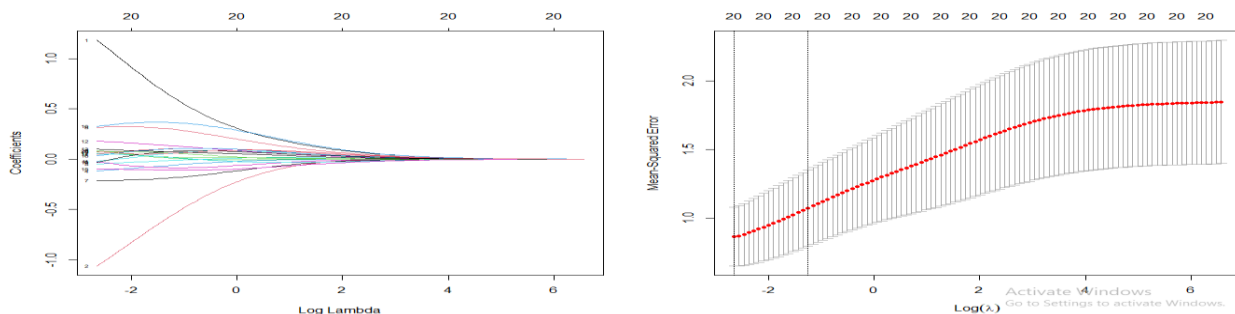
Table (10) analysis of variance

S.O.V	D.F	Sum Sq	Mean Sq	F value	Pr(>F)	R^2	Adjusted R^2
Model	20	75.59	3.7795				
Residuals	29	10.89	0.3755172	10.06478	3.010735e-08	0.8740	0.7872
Total	49	86.48					

From the table (10) We note from the value of (Pr(>F)) the significance of the model estimated using the method of ridge, as its value was close to zero and it is less than the level of significance ($\alpha = 0.05$) and thus rejected H_0.

We also note that the value of (R^2) is equal to (%87), which is interpreted as the ratio of the total change occurring in the response variable (number of children born) as a result of its effect on the values of the predictors variables (X) by the parameter (b) which links the mediation variable to the outcome variable in order to test the partial effect (C'), where the value of (R^2) can be considered as an indicator of the suitability of the estimated model or an indicator of the strength of the linear relationship between the number of children born and predictors variables.

We also note the value of ($Adjusted\ R^2$), which is equal to (%78), which is another measure to judge the extent to which the estimated model matches the data, assuming a change in the number of predictors variables. We can use ($Adjusted\ R^2$) to compare models that have different numbers of predictors variables.

Figure (5) a schematic diagram of the second model using the ridge method

Third Model: Estimating the Regression Model (M/X)

In this model (14), the effect of the parameters of the predicted variables (X) on the mediation variable (M) will be estimated in order to test the direct effect (a)

Null hypothesis: $H_0: a = 0$

and the table (11) below shows a summary of statistics for the estimated model.

Table (11) shows a summary of the estimated model statistics

Variable	Estimate	S.E.	t	P-Value
X_1	6.250e-01	2.840e-01	2.201	0.02773 *
X_2	8.019e-01	3.102e-01	2.585	0.00973 **
X_3	-7.287e-02	3.134e-01	-0.233	0.81611
X_4	-1.420e-01	3.135e-01	-0.453	0.65059
X_5	1.782e-01	3.077e-01	0.579	0.56244
X_6	-3.158e-01	3.156e-01	-1.001	0.31699
X_7	7.265e-02	3.161e-01	0.230	0.81819
X_8	-2.938e-01	3.266e-01	-0.900	0.36835
X_9	-2.692e-01	3.126e-01	-0.861	0.38907
X_{10}	1.731	3.122e-01	5.543	2.98e-08 ***
X_{11}	6.612e-02	3.178e-01	0.208	0.83517
X_{12}	2.022e-01	3.229e-01	0.626	0.53109
X_{13}	-5.785e-02	3.135e-01	-0.185	0.85359
X_{14}	3.416e-01	3.179e-01	1.075	0.28257
X_{15}	-4.555e-01	3.138e-01	-1.451	0.14667
X_{16}	-1.425e-01	3.139e-01	-0.454	0.64996
X_{17}	1.525e-01	3.176e-01	0.480	0.63107
X_{18}	6.039e-01	3.121e-01	1.935	0.05296
X_{19}	1.910	3.112e-01	6.138	8.36e-10 ***

Table (11) shows the estimated values of the model parameters (14) in addition to the standard errors and the test (t) and the values of (P-Value) where we notice that the variables (Age of the woman) (Woman's age at marriage) (Number of deceased children) (Gestational diabetes) have a

significant effect level of significance ($\alpha = 0.05$) and this is clear from the values of the column (P-Value) as well as we note from this The table shows that the rest of the variables do not have a significant effect in the study model, and this is evident from the values of (P-Value) in comparison with the level of significance ($\alpha = 0.05$).

To test the significance of the estimated model (14) by using the ridge method, we will test the hypothesis.

$H_0: a = 0$ and $H_1: a \neq 0$, where the table below (12) shows the analysis of variance table for the estimated model.

Table (12) analysis of variance

S.O.V	D.F	Sum Sq	Mean Sq	F value	Pr(>F)	R^2	Adjusted R^2
Model	19	6.3	0.3315789				
Residuals	30	15.88	0.5293333	0.6264086	0.8557959	0.7159	0.5360
Total	49	22.18					

From Table (12) we note that the value of (Pr (> F)) is not significant using the Ridge method as its value was not close to zero and it is greater than the level of significance (α = 0.05), thus accepting H_0.

We also note that the value of (R^2) is equal to (%71), which is interpreted the effect of the parameters of the predictors variables (X) on the mediation variable (M) which will be estimated in order to test the direct effect (a) , where the value of (R^2) can be considered as an indicator of the

suitability of the estimated model or an indicator of the strength of the linear relationship between the number of children born and predictors variables.

We also note the value of ($Adjusted\ R^2$), which is equal to (%53), which is another measure to judge the extent to which the estimated model matches the data, assuming a change in the number of predictors variables. We can use ($Adjusted\ R^2$) to compare models that have different numbers of predictors variables.
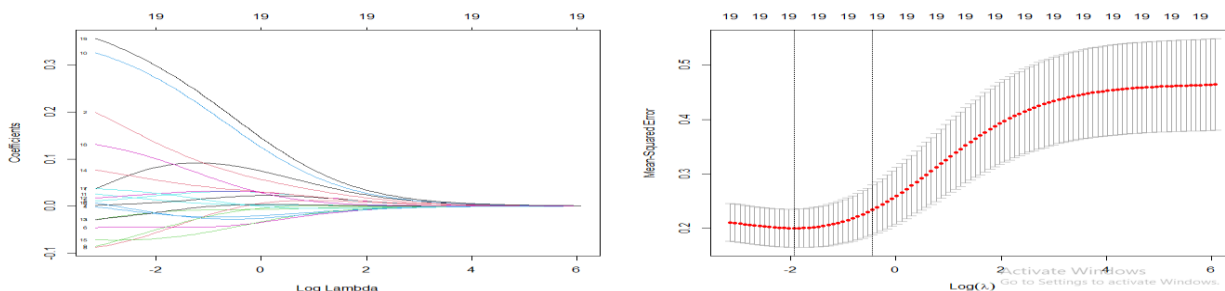
Figure (6) a schematic diagram of the third model using the ridge method

13. Conclusions

1- The effectiveness of the ridge regression method, which made the model more interpretable, in addition to choosing the easiest variables.

2- The emergence of the insignificance of some variables in the least squares method, the ridge regression method, and the lasso regression method despite its importance due to the presence of the problem of linear multiplicity and this is evident in the lack of significance of the mediation.

3- The emergence of some regression coefficients in the least squares method and the ridge regression method with an algebraic sign not identical to the reality of the studied phenomenon This is an indication of the inaccuracy of these methods in constructing a prediction model in the case of a multiplicity problem.

14. Recommendations

1- We recommend using the ridge regression method for selecting variables because it is able to reduce the parameter estimation to almost zero by imposing a restriction on the sum of the squares of the Rss regression.

2- We recommend using the Lasso method when the matrix $(X^{'} X)$ suffers from lack of an inverse value or when $P > n$. The Lasso method provides the best method for selecting the variables Variable Selection and is better than the ridge regression RR.

3- Not relying on the least squares method and the ridge regression method because they do not fully address the problem of linear multiplicity, especially in social studies, or when the number of independent variables is greater.

References

[1] Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173.

[2] Gentle, J.E. 2007. Matrix algebra: eory, computations, and applications in statistics. (Springer Texts in Statistics). New York, NY, USA: Springer.

[3] Grotta, A., & Bellocco, R. (2012). Causal mediation analysis on survival data: an application on the National March Cohort. PhD thesis, Univ. Milano-Bicocca, Milan.

[4] Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. Communication Monographs, 76(4), 408–420.

[5] Hoerl, A. E & Kennard, R. W (1970). "Ridge Regression: Biased Estimation for Non – Orthogonal Problems", Technometrics, Vol. 12, PP. 55-67.

[6] James, G., Witten, D., Hastie, T. and Tibshirani, R. 2013. An introduction to statistical learning: with applications in R. (Springer Texts in Statistics: 103). New York, NY, USA: Springer.

[7] MacKinnon, D. (2012). Introduction to statistical mediation analysis. Routledge.

[8] MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. Personality and Social Psychology Review, 19 (1), 30–43.

[9] MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007).

[10] MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. Annu. Rev. Psychol., 58, 593–614.

[11] MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research, 39(1), 99–128.

[12] Montgomery, D.C., Peck, E.A., and Vining, G.G. (2013). Introduction to Linear Regression Analysis. John Wiley & Sons, Inc., New Jersey, USA.

[13] Murtas, R. (2016). Mediation analysis for different types of Causal questions: Effect of Cause and Cause of Effect. Universita'degli Studi di Cagliari.

[14] Otter, T., Pachali, M. J., Mayer, S., & Landwehr, J. (2018). Causal inference using mediation analysis or instrumental variables-full mediation in the absence of conditional independence. *Available at SSRN 3135313*.

[15] Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40(3), 879–891.

[16] Tofighi, D., & Thoemmes, F. (2014). Single-level and multilevel mediation analysis. The Journal of Early Adolescence, 34 (1), 93–119.

[17] Wen, S. (2013). Estimation of multiple mediator model.