Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 21, Issue 4(2019) ISSN$_{Print}$: 1816-9171

Estimating The Effects of the Blood Sugar Model with Single Mediation Variable

Ahmad N. Flaih
Zain alabdin Aboud Kazem El-Husseini
Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq
Corresponding Author: Ahmad N. Flaih *Zain alabdin Aboud Kazem El-Husseini*

Abstract : *This paper presents a study of mediation analysis (with one moderator variable) and the role of this variable in moving the effect from the independent variable to the dependent variable. In this article, the delta method was used to estimate the standard error. Because this method violates some assumptions, especially with small sample sizes, the bootstrapping method was used to overcome such problems. According to the results obtained from the application of this model, the method of bootstrapping is better than the Delta method based on the results.*

Keywords: single mediator model; indirect effect; direct effect; causal inference

INTRODUCTION

The model of mediation analysis is of great importance in the knowledge and interpretation of the causes of the relationship between the cause variable and effect by knowing one or more of the variables of mediation. (Wen, 2013). It is known that the estimation of causal effects allows researchers to know whether the treatment affects the outcome, yet it does not tell us how this effect occurs.

Mediation analysis works to overcome such problem as it identifies the variables Interlaced (mediation variables) between independent variables and result, the causal mediation analysis formulated in the framework of Linear Structural Equation Modeling (LSEM) as in MacKinnon, 2008 (Rose, Chassin, Presson, & Sherman, 2000) (Imai, Keele, & Tingley, 2010).

Statistical methods of mediation analysis were discussed for several years in the psychological literature as early as 1948 where both talk Kenneth MacCorquodale and Paul E. Meehl about the value and the logical state of the variables of mediation (Wen, 2013).

Mediation has been assumed in its simplest form a third variable called a mediator (M) convey the effect from the independent variable (X) to the dependent variable (Y)($X \rightarrow M \rightarrow Y$) it is called indirect effect through the intermediary (M), either the direct effect is the effect of the independent variable (X) directly on the dependent variable Y (Y) (Murtas, 2016).

In the previous time, the variables of the mediation where called the processes variables (Judd & Kenny, 1981). where these variables describe the effect of the independent variable on a dependent variable, and these are called endpoints in the medical literature because they are variables that mediate between cause and effect (Prentice, 1989) (D. MacKinnon, 2012).

In overlapping research with causal inference and mediation analysis many points to an article (Baron & Kenny, 1986) basic analysis of mediation. Barron and Kenny are among the most widely consulted brochures in the research (Fairchild & McDaniel, 2017).

2- Single Mediation Model (SMM)

The only mediation model is one of the simplest models in mediation analysis. This model consists of three variables (the independent variable X, the dependent variable Y and the mediation variable M). This form is used when the independent variable affects a single moderator variable which causes an effect in the result variable. The Single mediator model has been applied which is also called individual mediation or simple mediation in many areas, including social, educational and behavioral sciences for the researcher's interest in knowing the side effects that is not studied (Wen, 2013).

In 1997 each of (Warner & Rountree, 1997) Studying mediation in social sciences use of poverty (independent variable) reduces the social links (the median variable), increasing the rates of theft and assault (dependent variable).

Figure (1) illustrates the relationship between three variables it is the simplest model of mediation effects where the shape represents a model for one intermediate variable, where the first variable is the independent variable (X), The second variable is the mediation variable (M), The third variable is the dependent variable (Y). In this figure there is a relationship with direct and indirect impact, the arrow between the independent variable (X) and the dependent variable (Y) indicates a direct correlation of the independent variable X to the Y result variable (X → Y). The second effect is the result of the effect of the independent variable X on the mediation variable M, which in turn affects the result variable Y (X → M → Y), which is called the indirect effect. The symbols (a, b, c) have a special significance where the symbol (a) refers to the effect of the independent variable X on the mediation variable M, the symbol (b) indicates the effect of the mediation variable M on the result variable Y, and the symbol C in figure (1) the total effect of the independent variable X on the result variable is Y (D. MacKinnon, 2012) (D. P. MacKinnon, Fairchild, & Fritz, 2007) (D. P. MacKinnon & Pirlott, 2015).

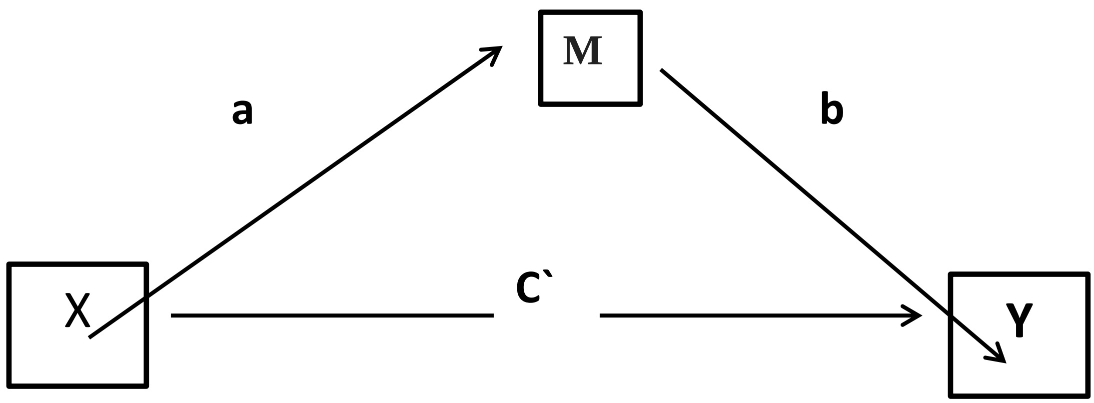


Figure (1): The indirect effect of the **X** variable on the **Y** variable is represented by the **M** mediation variable
Describe the model of single-mediation analysis with the following three regression equations:

$$\mathbf{M = i_1 + ax + e_1} \quad \ldots\ldots\ldots\ldots \mathbf{(1)}$$

$$\mathbf{Y = i_2 + Cx + e_2} \quad \ldots\ldots\ldots \mathbf{(2)}$$

$$\mathbf{Y = i_3 + C`x + bM + e_3} \quad \ldots\ldots\ldots\ldots \mathbf{(3)}$$

Where Y: represents the result variable (the dependent variable).
X: represents the independent variable.
M: represents the mediation variable or the overlapping variable.
C: is the relationship between the independent variable and the dependent variable in equation 2 and is called the total effect.
C`: represents the parameter that links the independent variable with the dependent variable under the influence of mediation in equation 3, which is called the partial effect.
a: is the parameter that conveys the effect of the independent variable on the mediation variable in equation 1.
b: is the parameter that conveys the effect of the mediation variable on the dependent variable in equation 3.
$\mathbf{i_1, i_2, i_3}$: are objections in each equation.
$\mathbf{e_1, e_2, e_3}$: Represents errors.

Equation 1 shows the model of the total effect in figure (2.3), and equations 2 and 3 define the mediation model in figure (2.3) (D. P. MacKinnon & Pirlott, 2015) (Otter, Pachali, Mayer, & Landwehr, 2018) (Tofighi & Thoemmes, 2014).

3- The problem of confounding

confounding is one from common problems facing many researchers. The problem of variation occurs when there is a mediation variable that links two or more variables in the mediation model and thus interprets the relationship between the variables (VanderWeele & Shpitser, 2013) (D. P. MacKinnon, Krull, & Lockwood, 2000).

There can be confounding about the relationship between variables (X → Y), (X → M), (Y → M) As shown in Fig (2.2). If any relationship of the former relationship is confounding and no adjustment is made to the confounding then the causal mechanism resulting from the analysis is a mirage. There are many confounding in psychological research, often psychological research on the effects of social strata and discrimination based on sex, environmental contexts, the history of alcohol use and other confounding variables. The confounding variables must be measured for a thorough investigation of the mediation processes (Valente, Pelham III, Smyth, & MacKinnon, 2017).

In a study of the analysis of mediation, we note that maternal smoking hurts infant mortality, especially in children with low birth weight (variable mediation), congenital defects in infants are positively associated with low birth weight as well as infant mortality,

At the same time, maternal smoking is directly associated with low birth weight and thus bias to the unmediated medium (VanderWeele, Mumford, & Schisterman, 2012).

The difference between mediation and confounding is that mediation is a chain between the cause and effect while the confusion affects both the cause and effect, confounding occurs prior to exposure such as prior medical history or demographic data available prior to exposure (Mascha, Dalton, Kurz, & Saager, 2013).

4- Effect of Mediation

One the reasons that motivate the researcher to know the effects process between of the three variables **(X, Y, M)** is to explain the causal process in which the independent variable **X** affects the dependent variable Y (James & Brett, 1984).

When studying any mediation process by which the relationship between the two variables **X**, **Y** is divided into two paths as explained in Fig (1.2), the first path is the direct effect in which the independent variable **X** affects the dependent variable **Y** (**X** → **Y**), the second track is where the independent variable **X** affects the variable **Y** by the mediation variable **M**, which is called indirect influence (**X** → **M** → **Y**).

The median effect is equal to the difference between the **C- C`** parameters. It is known that the overall effect is divided into direct effect represented by the parameter, **C**, and another indirectly effect represented by the parameters of (**a**, **b**) is being **ab** = **C - C`** estimate these parameters be equal except in cases where the size of the sample used in regression equations is different, for example, if the sample size used in equation 3 is different from the sample size used in both equations 2 and 1, in this case, the estimate of **C-C`** may be different or unequal to the estimation of the parameters **ab**. The parameters in the above three equations can be estimated using the regression method of the Ordinary Least Squares (OLS) or the Maximum Likelihood Estimation (MLE) method. After estimating the parameters, we can obtain a median estimate of **ab**, **C - C`** (D. P. MacKinnon et al., 2000) (Bareinboim & Pearl, 2012).

5- Product of Parameter Coefficients Testing

It a method used to test the importance of the mediation effect using the product and the standard error of the product the standard error derived from the Sobel equation was used in 1982 using the delta method based on the first-order tyler series approximation (Sobel, 1982).

$$S_{ab} = \sqrt{S_a^2\ b^2 + S_b^2\ a^2}$$

Where **a** and **b** represent the coefficients listed in the regression equations, $\boldsymbol{S_a^2}$ and $\boldsymbol{S_b^2}$ The standard error for **a** and **b** .

The standard errors that depend on the calculation are called the mediation effect ab a product of standard coefficients errors , and those that depend on C – C` are called the difference in the standard errors of the transactions Where c -c` is the standard error of the difference between regression coefficients which can be calculated as follow :

$$\mathbf{S}_{c\text{-}c`} = \sqrt{S_c^2 + S_{c`}^2 - 2\, r\, S_c\, S_{c`}}$$

after the estimation of standard errors, the upper and lower confidence limits of the effect of the mediator variable can be as follow :

$$\text{Upper confidence limit} = \text{mediated effect} + \mathbf{Z}\ (\mathbf{S_{ab}})$$

$$\text{Lower confidence limit} = \text{mediated effect} - \mathbf{Z}\ (\mathbf{S_{ab}})$$

The effects of standard error can be divided and compared them with **Z** distribution, for example, 1.96 and with a 95% confidence limit for large sample size, and have an impact estimate ab = c - c` the error of type I follows the **Z** statistic (D. P. MacKinnon & Luecken, 2011). What is mentioned is the confidence limits for the effect of the symmetric medium. as for the asymmetric confidence limits, they are more accurate because the effect of the mediator s natural distribution is not required may be another distribution. asymmetric confidence limits can be calculated by means of reconfiguration or use of critical values of product distribution (D. MacKinnon, 2012).

6-Bootstrapping

In many situations the sample size is not enough to study, this is one of the problems facing researchers in many studies and to overcome this problem, Bolger & Shrout suggested in 2002, the use of Bootstrapping method. the delta method of estimating the standard error does not work well when the sample size is not sufficiently large. This limitation leads to the use of the Bootstrapping method.

Bootstrapping can be used to find the standard errors of the estimated parameter when the sampling distribution of the estimated parameter is unknown, as well as Bootstrapping, can be used to assess the accuracy of the estimated model and to assess its prediction accuracy (Wen, 2013).

The Bootstrapping method is one of the common ways of estimating indirect effects. Bootstrapping depends on the restructuring with the replacement of a large number of times. Meaning that if we have a sample of size n the bootstrapping process is performed by taking **K** of the samples with the repetition and the replacement process from the original sample size and preferably K = 1000 at least (Preacher & Hayes, 2008).

Each sample has its characteristics, such as the medium as well as estimating the indirect effects of each sample. This method is used to conduct sample distributions as a basis for confidence intervals and hypothesis testing (D. MacKinnon, 2012).

7- Mediation Analysis Steps for Single Mediation Model

A set of steps were developed by both (Baron & Kenny, 1986) (James & Brett, 1984) (Judd & Kenny, 1981) for mediation analysis:

Step 1: There should be an effect of the independent variable (**X**) on the dependent variable (**Y**) by parameter **C** as shown in Equation 2.

Step 2: The independent variable (**X**) must have an effect on the mediation variable (**M**) by the parameter **a** an as shown in equation 1.

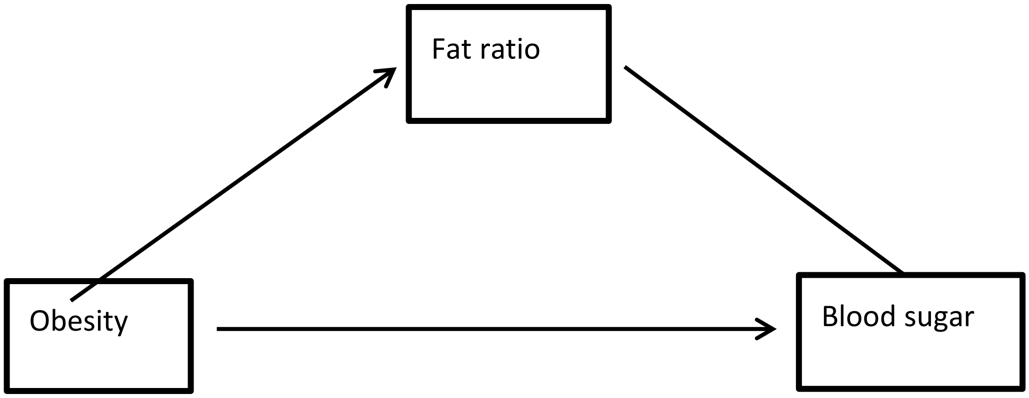
Step 3: The mediation variable (**M**) should affect the dependent variable (**Y**) by parameter **b** as shown in equation 3.

Step 4: To prove that the variable (**M**) intermediate both the independent variable (**X**) and the dependent variable (**Y**) fully the effect of the independent variable on the dependent variable (direct effect) of parameter **C`** in equation 3 should be completely unimportant.

When the four steps are fully implemented, we can say that the data conforms to the hypothesis that the **M** variable is fully averaging the independent variable **X** and the dependent variable **Y**, but in the failure to meet step four, in this case, there is partial mediation.

A study conducted by (D. P. MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) explained that the most important steps of mediation are step 2 and step 3, ie, the coefficient a is statistically significant in equation 1 and also the coefficient b in equation 3.

In this model data on obese people were used to directly influence lipids and their effect on blood sugar indirectly. The data were collected by monitoring the patient records at the diabetes center at Diwaniyah teaching hospital. A sample of 140 patients was obtained in 2018. Obesity was used as a independent variable (**X**), fat percentage as a variable of mediation (**M**), and diabetes as a variable (**Y**).



Figure(2): path way of ratio the obesity with the fat ratio with the blood sugar.

8- Data Analysis Results

8-1 Mediation Analysis of Obesity Data

- A- Causal procedure strategy to Establish mediation

First, we check the strength between the effect of obesity (X) and the ratio of fat (M) and blood sugar (Y) by tracking the relationship path using the correlation matrix as shown in the bellows :

	M	X	Y
M	1	0.0497	0.0415
X	0.0497	1	0.1163
Y	0.04158	0.1163	1

Table (1): Correlation matrix among independent variable , dependent variable, and mediators.

The correlation matrix shown in table (1) shows a correlation between obesity and fat in the rate of (0.0497), as well as a correlation between obesity and blood sugar in the rate of (0.1163). There is also a correlation between fat and blood sugar ratio in the rate of (0.0415).

-B- Product of Coefficient Approach

When using lm regression method results in R programs, we have the following estimates of the proposed model :

$Y = i_1 + CX + e_1$ ………(1)

The equation (1) shows the total (direct) effect of obesity on blood sugar.

Parameters	$\hat{C}$
Estimate	2.48869
Std. Error	0.04914
T value	50.64
Pr(>|t|)	<2e-16
C.I	2.561- 2.400

Residual standard error	56.25
Multiple R-squared	0.9486
Adjusted R-squared	0.9482
F-statistic	2565
p-value	< 2.2e-16

Table (2): Estimates total effect (direct effect) $\hat{C}$.

From table (2) we observe the total effect estimate (Direct effect) of obesity on the blood sugar level at a rate of (2.48869) and a standard error at a rate of (0.04914) and there is a significant effect $p< 0.05$.

The total error in the model was estimated at a rate of (56.25), and with a coefficient of determination at a rate of (0.9486), ie that it is possible to determine at a rate of (0.9486) of the change in the sugar level depending on obesity, while the significance of the model in general is ($p<2.2e-16$) This indicates a significant model.

From table (2) we obtain the following estimations:

$$\text{Total effect } (\hat{C}_{\text{direct effect}}) = 2.481$$

$$\text{Std. Error } (S_{\hat{C}}) = 0.041$$

$$\text{Upper Confidences Limits (UCL)} = \hat{C}_{\text{direct effect}} + Z^*S_{\hat{C}}$$

$$= 2.481 + 1.96\,(0.041)$$

$$= 2.561$$

$$\text{Lower Confidences Limits (LCL)} = \hat{C}_{\text{direct effect}} - Z^*S_{\hat{C}} \qquad = 2.400$$

$$Y = i_2 + C`X + bM + e_2 \qquad \ldots\ldots\ldots\ldots (2)$$

Equation (2) represents the effect of obesity on blood sugar (partial effect) with another indirect effect (mediation effect) represented by fat ratio .

Parameters	$\hat{C}`$	$\hat{b}$
Estimate	1.74969	0.26457
Std. Error	0.20586	0.07175
T value	8.499	3.687
Pr(>|t|)	2.76e-14	0.000325
C.I	2.153-1.346	0.4052- 0.123

Residual standard error	53.86
Multiple R-squared	0.9532
Adjusted R-squared	0.9525
F-statistic	1405
p-value	< 2.2e-16

Table (3): Estimates the Partial effect $\hat{C}`$ and the parameter $\hat{b}$.

From table (3) we observe the partial effect estimate of obesity on the blood sugar level (1.74969) and the standard error at a rate of (0.20586) and estimate the effect of fat percentage (effect of mediation) by (0.26457) and a the standard error at a rate of (0.07175) there is a significant effect where $p<0.05$.

The total error in the model was estimated at a rate of (53.86), and with a coefficient of determination at a rate of (0.9532), ie that it is possible to determine at (0.9532) of the change in the sugar level depending on obesity and Fat ratio, while the significance of the model in general is ($p< 2.2e-16$) this indicates a significant model.

From table (3) we obtain the following estimations:

$$\text{Partial effect } (\hat{C}`) = 1.74969$$
$$\text{Std. Error } (S_{\hat{C}`}) = 0.20586$$
$$\text{Upper Confidences Limits (UCL)} = \text{Partial effect } (\hat{C}`) + Z^*S_{\hat{C}`}$$
$$= 1.74969 + 1.96\ (0.20586) = 2.153$$

$$\text{Lower Confidences Limits (LCL)} = \text{Partial effect } (\hat{C}`) - Z^*S_{\hat{C}`} = 1.346$$
$$\hat{b}_{\text{indirect effect}} = 0.26457$$
$$\text{Std. Error } (S_{\hat{b}}) = 0.07175$$
$$\text{Upper Confidences Limits (UCL)} = \hat{b}_{\text{indirect effect}} + Z^*S_{\hat{b}}$$
$$= 0.26457 + 1.96\ (0.07175)$$
$$= 0.4052$$
$$\text{Lower Confidences Limits (LCL)} = \hat{b}_{\text{indirect effect}} - Z^*S_{\hat{b}}$$
$$= 0.123$$

$$M = i_3 + aX + e_3 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3)$$

Equation (3) represents the effect of obesity on fat ratio In a manner direct.

Parameters	$\hat{a}$
Estimate	2.79321
Std. Error	0.05563
T value	50.21
Pr(>|t|)	<2e-16
C.I	2.902-2.684

Residual standard error	63.67
Multiple R-squared	0.9478
Adjusted R-squared	0.9474
F-statistic	2521
p-value	< 2.2e-16

Table (4): Estimates parameter $\hat{a}$.

From table (4) we observe the total effect estimate of obesity In a manner direct on fat percentage at a rate of (2.79321) and a standard error at a rate of (0.05563) there is a significant effect where $p< 0.05$.

The total error in the model was estimated at a rate of (63.67), and with a coefficient of determination at a rate of (0.9478), ie that it is possible to determine (0.9478) of the change in the fat percentage depending on obesity, while the significance of the model in general is ($p < 2.2e{-16}$) This indicates a significant model.
From table (4) we obtain the following estimations:

$$\hat{a}_{\text{direct effect}} = 2.79321$$

$$\text{Std. Error } (S_{\hat{a}}) = 0.05563$$

$$\text{Upper Confidences Limits (UCL)} = \hat{a}_{\text{direct effect}} + Z^* S_{\hat{a}}$$

$$= 2.79321 + 1.96\,(0.05563) = 2.902$$

$$\text{Lower Confidences Limits (LCL)} = \hat{a}_{\text{direct effect}} - Z^* S_{\hat{a}}$$

$$= 2.684$$

The standard error can also be calculated:

$$\hat{a}_{\text{direct effect}} * \hat{b}_{\text{indirect effect}} = 2.79321 * 0.26457 = 0.7389996$$

Obesity **(X)** was significantly associated with blood sugar **(Y)**
($\hat{c}_{\text{direct effect}} = 2.48869$, $S_{\hat{c}} = 0.04914$, $t_{\hat{c}} = 50.64$), providing evidence of a statistically significant intervention effect at 2.48869 units. There was a statistically significant intervention effect on fat effect **(M)** ($\hat{a}_{\text{direct effect}} = 2.79321$, $S_{\hat{a}} = 0.05563$, $t_{\hat{a}} = 50.21$). The intervention increased fat by 2.79321 after the intervention. The effect of blood sugar mediator was statistically significant ($\hat{b}_{\text{indirect effect}} = 0.26457$, $S_{\hat{b}} = 0.07175$, $t_{\hat{b}} = 3.687$).
The effect of obesity was associated with a change in the proportion of fat, which led to a change in blood sugar. If the adjusted effect of the intervention is statistically significant ($\hat{C}'_{\text{partail effect}} = 1.74969$, $S_{\hat{C}'} = 0.20586$, $t_{\hat{C}'} = 8.499$). The median influence rating is equal $\hat{a}_{\text{indirect effect}} * \hat{b}_{\text{indirect effect}} = (2.79321)\ (0.26457) = \hat{C}_{\text{direct effect}} - \hat{C}'_{\text{partial effect}} = 2.48869 - 1.74969 = 0.739$. The mean effect of obesity through a change in fat ratio was 0.739 units of change in blood sugar. Using the standard error equation to estimate the median effect as shown in the following:

$$S_{\hat{a}_{\text{indirect effect}} * \hat{b}_{\text{indirect effect}}} = \sqrt{s_{\hat{a}}^2\, \hat{b}^2 + \hat{a}^2\, s_{\hat{b}}^2}$$

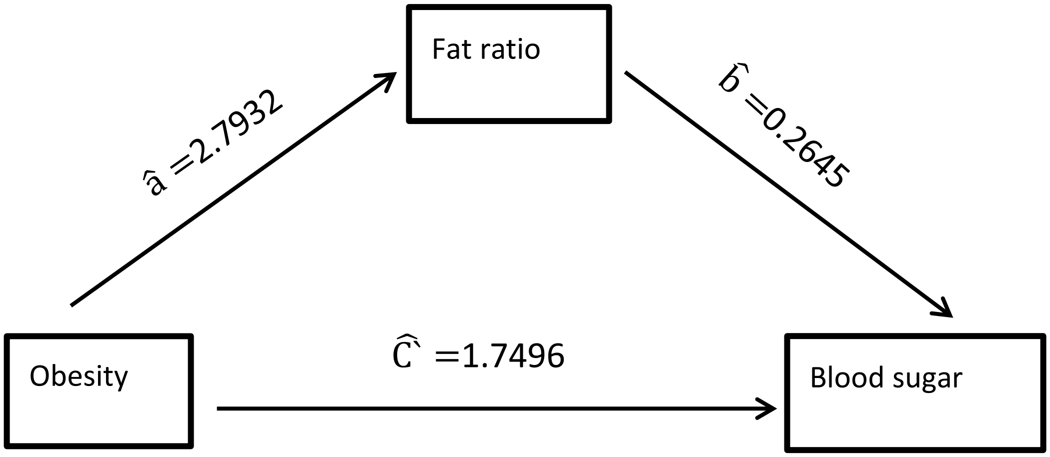
$$= \sqrt{(0.05563)^2\,(0.26457)^2 + (2.79321)^2\,(0.07175)^2}$$

$$= 0.2009525$$

The confidence limits were as follows:

$$\text{Upper Confidences Limits (UCL)} = \text{Mediated effect} + Z^* S_{\hat{a}\hat{b}}$$
$$= 0.739 + 1.96(0.2009525)$$
$$= 1.132867$$
$$\text{Lower Confidences Limits (LCL)} = \text{Mediated effect} - Z^* S_{\hat{a}\hat{b}}$$
$$= 0.739 - 1.96(0.2009525)$$
$$= 0.3451331$$

The confidence limits were as follows all the results show that there are a statistical significance and an indication of the effect of the three variables (obesity, lipids and blood sugar).



Figure(3): Shows estimates of variables on the chart.

- C - Bootstrapping Estimation

Either when using the bootstrapping method to estimate the intervals of confidence as follows :

To obtain more accuracy, in bootstrapping, I chose to take 5,000 samples of 1500 cases with replacement from the original sample and calculated each mediated effect. The appendix includes the R code with sample command of bootstrapping.

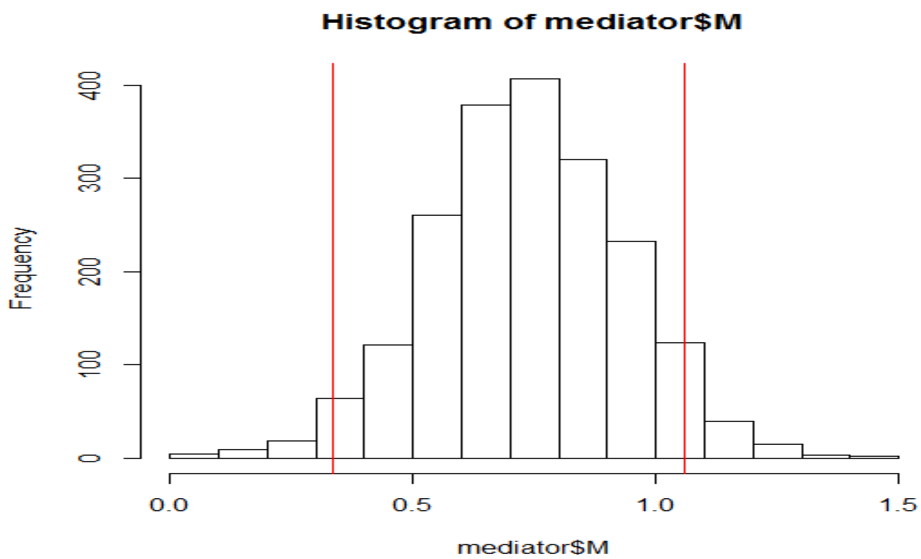


Figure (4): The effect diagram is shown by blood sugar. Red lines represent the minimum and the highest for every 95% of the confidence interval.

The above diagram shows the confidence interval for the mediation variable (Fat ratio) with a confidence interval of 95%. Since the period is limited between (0.3 - 1.1), I,e, zero does not belong to this period, it can be said that the effect of the mediation variable is different from zero and therefore there is an effect for this variable.

9- Conclusions

We can conclude from the results obtained from the analysis of the data of the Single mediation model of obesity in the previous chapter the following:

1- The effect of the independent variable (obesity) on the dependent variable (blood sugar) (total effect) is significant and at a rate of (2.48869).

2- The effect of the independent variable (obesity) on the dependent variable (fat percentage) (partial effect), with a presence of mediation variable () is significant and at a rate of (1.74969).

3- The effect of the mediation variable (fat percentage) on the dependent variable (blood sugar) is significant and at a rate of (0.2645).

By observing the correlation between the studied variables and the estimation of the parameters, as well as the estimated confidence intervals, it turns out that the method of bootstrapping is better than multiplying the coefficients as well as the steps of mediation analysis.

where bootstrapping is based on resampling the non sufficient large sample which makes different values of the estimates for the parameter of interest ,the we get a range (interval) of estimates rather than one point estimate . therefore bootstrapping technique is preferable on the delta method ,as well as, the steps of mediation analysis.

Reference

Bareinboim, E., & Pearl, J. (2012). Causal inference by surrogate experiments: z-identifiability. *ArXiv Preprint ArXiv:1210.4842*.

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.

Fairchild, A. J., & McDaniel, H. L. (2017). Best (but oft-forgotten) practices: a mediation analysis. *The American Journal of Clinical Nutrition*, *105*(6), 1259–1271.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*(2), 307.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*(5), 602–619.

MacKinnon, D. (2012). *Introduction to statistical mediation analysis*. Routledge.

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

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83.

MacKinnon, D. P., & Luecken, L. J. (2011). Statistical analysis for identifying mediating variables in public health dentistry interventions. *Journal of Public Health Dentistry*, *71*, S37–S46.

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.

Mascha, E. J., Dalton, J. E., Kurz, A., & Saager, L. (2013). Understanding the mechanism: Mediation analysis in randomized and nonrandomized studies. *Anesthesia & Analgesia*, *117*(4), 980–994.

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

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.

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.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine*, *8*(4), 431–440.

Rose, J. S., Chassin, L., Presson, C. C., & Sherman, S. J. (2000). Contrasts in Multiple M ediator Models. In *Multivariate applications in substance use research* (pp. 155–174). Psychology Press.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290–312.

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

Valente, M. J., Pelham III, W. E., Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: What it is and how to address it. *Journal of Counseling Psychology*, *64*(6), 659.

VanderWeele, T. J., Mumford, S. L., & Schisterman, E. F. (2012). Conditioning on intermediates in perinatal epidemiology. *Epidemiology (Cambridge, Mass.)*, *23*(1), 1.

VanderWeele, T. J., & Shpitser, I. (2013). On the definition of a confounder. *Annals of Statistics*, *41*(1), 196.

Warner, B. D., & Rountree, P. W. (1997). Local social ties in a community and crime model: Questioning the systemic nature of informal social control. *Social Problems*, *44*(4), 520–536.

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