ISSN NO: 2769-996X
Date of Publication: 13-04-2022

Using the R statistical programming language to estimate the mediation variables Influencing the presence of the proportion of elements in the soil

Zainalabideen AL-Husseini
Al-Mustaqbal University College, Babylon, Hilla, Iraq
zainalabden.aboad@mustaqbal-college.edu.iq
Hazim mohammed Dayekh
Al-Mustaqbal University College, Babylon, Hilla, Iraq
hazem.mohammed@mustaqbal-college.edu.iq

Abstract: The research aims to study the effect of population momentum and temperatures on the proportion of elements in the soil, as multiple mediation models were applied with regression models to study and indicate the effect of the variables under study. Those samples were laboratory to know the proportions of the effect. A group of elements were studied, including (Na, Mg, K, CI, Ca), and after collecting and tabulating the data, the data was analyzed using the statistical program (R Program).

Keywords: Multiple mediator model; Effects Mediation; direct effect; causal inference

Introduction

Most of the research in various studies focuses on generating a relationship between two variables, namely, the cause and effect variable, while there are variables that have a clear impact and are influencing the results called mediation variables. When adding the mediating variable (Z) to the analysis model, The matter may be more complicated because adding the new variable and considering it as a mediator between the cause and effect variables, where the mediation variable (Z) is indirectly affecting the result in addition to the direct effect produced by the cause variable, $X \rightarrow Z \rightarrow Y$ (Grotta & Bellocco, 2012).
In this paper, we will study a special case of mediation models related to the multiple mediation model where we study the simplest multiple mediation model with an applied case. It is known that the simplest multiple mediation models consist of two intermediate variables that perform their function to transfer the effect between each of the cause variable and the result variable, as shown in the following figure.

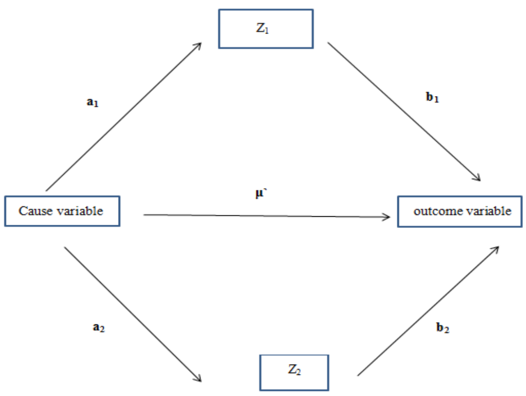


Figure 1 : Diagram representing the effects of variables .

Multi-mediator Model

Multiple mediation model

When starting to study the relationship between two variables, it is obvious that is a set of overlapping and explicit variables that indirectly affect the results. Many researchers are interested in such

ISSN NO: 2769-996X
Date of Publication: 13-04-2022

variables and are the focus of their study, but a number of these variables are often neglected and may be of im portance to alter the final results of the study. (Briggs, 2006), In the event that there is a set of variables, or at least two variables, it is considered as a case of mediation. Here, the researcher is required to rely on the multiple mediation model to reach the best results. According to several studies, we note the limited use of the mentioned model, due to the ambiguity of the analytical methods in addition to the difficulty of the multiple mediation models. (Cheung, 2007) (D. P. MacKinnon, 2000).

Estimating mediation variables using the regression equation

The following models represent equations for mediation influences:

$$Y = \Upsilon_1 + \mu X + e_1 \ldots\ldots\ldots\ldots 1$$

$$Y = \Upsilon_2 + \mu` X + b_1 Z_1 + b_2 Z_2 + e_2 \ldots\ldots\ldots 2$$

$$Z_1 = \Upsilon_3 + a_1 X + e_3 \ldots\ldots\ldots\ldots\ldots 3$$

$$Z_2 = \Upsilon_4 + a_2 X + e_4 \ldots\ldots\ldots 4$$

Were

Y: is the result variable.

X: indicates the cause variable

.Z1, Z2: will be used as symbols for the mediation variables

.µ`: the parameter that is under the influence of the mediation variables

.µ: the parameter that is in the absence of mediation variables

. a1: represents the parameter connecting X and Z1

. a2: represents the parameter connecting X and Z2

.b1: The parameter that links the first mediation variable to the outcome variable

.b2: The parameter that connects Z2 and Y

e1, e2, e3: represents errors (D. MacKinnon, 2012).

In the multi-mediated model as in Figure (1), the total effect should be equal to the sum of the direct effect of coefficient µ` and the indirect effect of Z1 and Z2 (Preacher & Hayes, 2008).

Total effect = $\mu` + a_1 b_1 + a_2 b_2$

Mediation Effects

The overall effect can be divided into direct effects of the coefficient µ ` and another is mediation. That is, `. $a_1 b_1 + a_2 b_2 = \mu - \mu`$. (Hayes, 2009).

Confidence intervals and significance testing for the effects of indirect variables

Estimating the standard error and comparing the resulting Z results with the critical value of the standard normal distribution is one of the most used methods to test the significance of the indirect effect (D. P. MacKinnon, Lockwood, & Williams, 2004).

Confidence intervals (CI) can also be calculated depending on the variables of mediation and standard error, and it is well known that (CI) used standard error of estimation, and for this reason, We note that (CI) may have been used to provide the largest number of effect values. Confidence intervals are commonly used in research because they require the researcher to take into account the value of the effect in addition to its statistical significance (Harlow, n.d.). To test the significance of the influence of indirect variables, we need to calculate the standard error of the sample by the median, and the tests commonly used to estimate the standard errors of the effects of mediation variables are as follows:

Steps to Establish Mediation

:There are a set of basic steps for indirect effects that are commonly used in research

Step 1: There must be an effect of the cause variable X on the outcome variable Y through the parameter µ`

Step 2: There must be an effect of the cause variable X on mediators Z1 and Z2.

Step 3: There must be mediation variables on the dependent variable Y after controlling for the cause variable X with factors b1, b2.

ISSN NO: 2769-996X
Date of Publication: 13-04-2022

Step 4: There must be an effect of the cause variable X on the dependent variable Y (direct effect μ`) to achieve complete mediation (Wen, 2013).

Product of Parameter Coefficients Testing

The standard error formula for the internal effect of the multiple mediation the standard error is given as below :

$$S_{a1b1} = \sqrt{S_{a1}^2 \; b_1^2 + S_{b1}^2 \; a_1^2}$$

Other formulas for standard error can be used for a multiple mediation :

$$S_{a1b1} + S_{a2b2} = \sqrt{S_{a1}^2 \; b_1^2 + S_{b1}^2 \; a_1^2 + S_{a2}^2 \; b_2^2 + S_{b2}^2 \; a_2^2 + 2a_1 \; a_2 \; S_{b1b2}}$$

The equation above can be rewritten as follows :

$$S_{a1b1} + S_{a2b2} = \sqrt{S_{a1b1}^2 + S_{a2b2}^2 + 2a_1a_2 \; S_{b1b2}}$$

In the case of the total effect, the standard error is as follows: $S_{\mu-\mu`} = \sqrt{S_{\mu}^2 + S_{\mu`}^2 - 2r \; S_{\mu}S_{\mu`}}$

Estimation of the standard error and median effect can be used to construct confidence limits as in equations (2-3-7-2) for similar periods (D. MacKinnon, 2012)

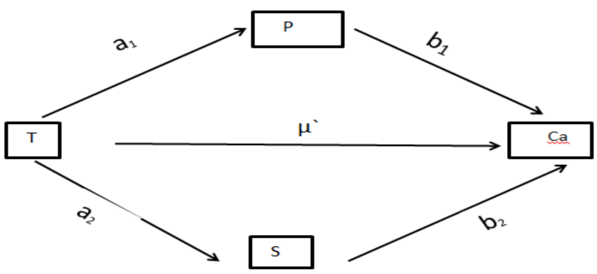
Bootstrapping

In many studies, the sample size is not sufficient for the study, and this is one of the most important problems that the researcher suffers from. While bolger & shrout suggested in 2002 using Bootstrapping method to overcome such problems. It is known that the delta method for estimating the standard error does not work well when the sample size is insufficient, and for this reason the researcher resorts to using the bootstrapping method (Preacher & Hayes, 2008).

Bootstrapping is one of the most popular methods for estimating mediation variables. Bootstrapping is based on restructuring with replacement as many times as possible. In other words, if we have a sample of size n, the bootstrapping process is performed by taking R samples with iteration and the substitution process from the original sample size preferably R = 1000 at least (Wen, 2013).

In such division, each sample has its own characteristics, such as measures of propensity as well as dispersion, as well as estimation of the indirect effects of each sample. This method is used to perform sample distributions as a basis for confidence intervals and to test hypotheses (**Keeny,2018).**

In this paper we deal with a study of the multiple mediation model based on data for a case study represented by knowing the effect of temperature T (the cause variable) on the type of soil S and population inflation P (mediating variables), which in turn affect the proportion of elements in the soil where the study was conducted on the element calcium Ca (result variable).



Figure(2): the pathway between the effect of temperature T and soil S and population inflation P and element calcium Ca .

Data analysis using programming language R

A- <u>Product of Parameter Coefficients Testing</u>

Relying on regression models, we get the following estimates

$$\mathbf{Y = \Upsilon_1 + \mu X + e_1 \ldots\ldots\ldots\ldots 1}$$

Equation 1 shows the total effect for the reason variable (Temperatures) on the calcium element.

Parameters	$\hat{\mu}$
Estimate	4.08
Std. Error	0.27
T value	19.70
Pr(>|t|)	<2e-16
C.I	4.479-3.67

Residual standard error	14.10
Multiple R-squared	0.901
Adjusted R-squared	0.906
F-statistic	389.3
p-value	< 2.2e-16

Table(1): Estimates total effect (Direct effect) $\hat{\mu}$.

From table (1) we observe the total effect estimate of the reason variable (Temperatures) on the calcium element at a rate of (4.08) and a standard error at a rate of (0.27) and there is a significant effect $p<0.05$.
The total error in the model was estimated at (14.10), and with a coefficient of determination of (0.901), ie that it is possible to determine (0.9089) of the change in the reason variable (Temperatures) on the calcium element , while the significance of the model in general is ($p< 2.2e-16$) This indicates a significant model.
From table (1) we obtain the following estimations:

$$\text{Total effect } (= \hat{\mu}_{\text{direct effect}}) = 4.074$$

$$\text{Std. Error } (S_{\hat{\mu}}) = 0.206$$

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

$$= 4.0748 + 1.96\,(0.2065) = 4.47$$

$$\text{Lower Confidences Limits (LCL)} = \hat{\mu}_{\text{direct effect}} - Z^* S_{\hat{\mu}} = 3.670$$

$$\mathbf{Y = \Upsilon_2 + \mu' X + b_1 Z_1 + b_2 Z_2 + e_2 \ldots\ldots\ldots 2}$$

Equation (2) represents the effect of the reason variable (Temperatures) on the calcium element (partial effect) with another Mediation effect represented by (**Z_1**) and (**Z_2**).

Parameters	$\hat{\mu}'$	$\hat{b}_1$	$\hat{b}_2$
Estimate	1.628	0.864	10.143
Std. Error	0.316	0.124	3.467
T value	5.15	6.96	2.92
Pr(>|t|)	8.79e-06	3.14e-08	0.0058
C.I	2.248-1.009	1.107-0.62	16.94-3.35

Residual standard error	8.43
Multiple R-squared	0.969
Adjusted R-squared	0.966
F-statistic	387.9
p-value	< 2.2e-16

Table (2): Estimates Partial effect and The parameters $\hat{b}_1$ and $\hat{b}_2$.

From table (2) we observe the partial effect estimate of the effect of the reason variable (Temperatures) on the calcium element at a rate of (1.628) , the standard error at a rate of (0.316) and estimate the effect of z1 (effect of mediation 1) at a rate of by (0.864) , the standard error at a rate of (0.124) and estimate the effect of the z2 (effect of mediation 2) at a rate of by (10.143) , the standard error at a rate of (3.467) , there is a significant effect where $p<0.05$.

The total error in the model was estimated at (8.43), and with a coefficient of determination of (0.9692), ie that it is possible to determine (0.9692) of the change in the calcium element depending on the z1 and z2 , 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:

Partial effect ($\hat{\mu}_{\text{partial effect}}$) = 1.628

Std. Error ($S_{\hat{\mu}}$) = 0.316

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

= 1.628+ 1.96 (0.316) = 2.248

Lower confidences limits (LCL) = $\hat{\mu}_{\text{partial effect}} - Z^*S_{\hat{\mu}}$

1.009=

$\hat{b}_{1\text{ indirect effect}}$ = 0.864

Std. Error ($S_{\hat{b}1}$) = 0.124

Upper Confidences Limits (UCL) = $\hat{b}_{1\text{ indirect effect}} + Z^*S_{\hat{b}1}$

= 0.864 + 1.96 (0.124)

= 1.107

Lower Confidences Limits (LCL) = $\hat{b}_{1\text{ indirect effect}} - Z^*S_{\hat{b}1}$

= 0.621

$\hat{b}_{2\text{ indirect effect}}$ = 10.143

Std. Error ($S_{\hat{b}2}$) = 3.467

Upper Confidences Limits (UCL) = $\hat{b}_{2\text{ indirect effect}} + Z^*S_{\hat{b}2}$

= 10.143 + 1.96 (3.467)

=16.938

Lower Confidences Limits (LCL) = $\hat{b}_{2\text{ indirect effect}} - Z^*S_{\hat{b}2}$

= 3.347

$$Z_1 = \Upsilon_3 + a_1X + e_3 \quad ....3$$

Equation (3) represents the effect of temperature T on population inflation P in a manner direct .

Parameters	$\hat{a}_1$
Estimate	2.124
Std. Error	0.164
T value	12.9
Pr(>|t|)	1.19e-15
C.I	2.44-1.80

ISSN NO: 2769-996X
Date of Publication: 13-04-2022

Residual standard error	11.27
Multiple R-squared	0.810
Adjusted R-squared	0.805
F-statistic	166.4
p-value	<1.193e-15

Table (3): Estimates parameter $\hat{a}_1$.

From table (3.14) we observe the total effect estimate of temperature In a manner direct on population inflation at a rate of (2.1244) and a the standard error at a rate of (0.1647) , there is a significant effect where $p<0.05$.

The total error in the model was estimated at (11.27), and with a coefficient of determination of (0.810), Ie that it is possible to determine (0.810) of the change in the population inflation depending on the temperature, while the significance of the model in general is ($p< 1.193e-15$) this indicates a significant model.

From table (3-14) we obtain the following estimations:

$$\hat{a}_{1\ \text{direct effect}} = 2.124$$

$$\text{Std. Error } (S_{\hat{a}1}) = 0.164$$

$$\text{Upper Confidences Limits (UCL)} = \hat{a}_{1\ \text{direct effect}} + Z^*S_{\hat{a}1}$$

$$= 2.124 + 1.96\,(0.164)$$

$$= 2.44$$

$$\text{Lower Confidences Limits (LCL)} = \hat{a}_{1\ \text{direct effect}} - Z^*S_{\hat{a}1}$$

$$= 1.80$$

$$\mathbf{Z_2 = \Upsilon_4 + a_2X + e_4} \ \ldots\ldots\ldots\ \mathbf{4}$$

Equation (4) represents the effect of the temperature T on soil S in a manner direct .

Parameters	$\hat{a}_2$
Estimate	0.06008
Std. Error	0.00589
T value	10.19
Pr(>|t|)	1.48e-12
C.I	0.0713-0.04

Residual standard error	0.4033
Multiple R-squared	0.7271
Adjusted R-squared	0.7201
F-statistic	103.9
p-value	< 1.484e-12

Table (4): Estimates parameter $\hat{a}_2$.

From table (4) we observe the total effect estimate of the temperature In a manner direct on the soil at a rate of (0.06008) , the standard error at a rate of (0.00589) there is a significant effect where $p<0.05$.

 ISSN NO: 2769-996X
Date of Publication: 13-04-2022

The total error in the model was estimated at (0.4033), and with a coefficient of determination of (0.727), ie that it is possible to determine (0.7271) of the change in the soil depending on the temperature , while the significance of the model in general is (p<1.484e-12) this indicates a significant model.

From table (4) we obtain the following estimations:

$\hat{a}_{2\ \text{indirect effect}} = 0.06008$

Std. Error ($S_{\hat{a}2}$) = 0.00589

Upper Confidences Limits (UCL) = $\hat{a}_{2\ \text{direct effect}} + Z^*S_{\hat{a}2}$

= 0.060 + 1.96 (0.0058) = 0.0713

Lower Confidences Limits (LCL) = $\hat{a}_{2\ \text{direct effect}} - Z^*S_{\hat{a}2}$

= 0.0486

Standard error can also be calculated:

$\hat{a}_{1\text{indirect effect}} * \hat{b}_{1\ \text{indirect effect}} = 2.1244 * 0.8646$

= 1.836756

$\hat{a}_{2\ \text{indirect effect}} * \hat{b}_{2\ \text{indirect effect}} = 0.06008*10.143$

= 0.609391

It was found that the temperature significantly affected calcium element (**Y**) ($\hat{C}_{\text{direct effect}}$ = 4.0748, Std. Error ($S_{\hat{c}}$) = 0.2065, $t_{\hat{c}}$ = 19.73), providing evidence of a statistically significant intervention effect at 4.0748 units. The effect of the temperature was statistically significant for both z1 ($\hat{a}_{1\ \text{indirect effect}}$ = 2.124, Std. Error ($S_{\hat{a}1}$) = 0.164, and $t_{\hat{a}1}$ = 12.9). As well as on z2
($t_{\hat{a}2}$ = 10.19, $\hat{a}_{2\ \text{indirect effect}}$ = 0.06008, Std. Error ($S_{\hat{a}2}$) = 0.005894).
The effect of z1 was also statistically significant
($\hat{b}_{1\ \text{indirect effect}}$ = 0.8646, Std. Error ($S_{\hat{b}1}$) = 0.124, $t_{\hat{b}1}$ = 6.967), The results of z2 also showed a statistically significant effect ($\hat{b}_{2\ \text{indirect effect}}$ = 10.143, Std. Error ($S_{\hat{b}2}$) = 3.4672, $t_{\hat{b}2}$ = 2.925).
the temperature has had a difference in the proportion of z1 as well as a change in the proportion of z2 , leading to a change in calcium element. Where the effect was statistically significant
(Partial effect ($\hat{\mu}_{\text{partial effect}}$) = 1.6287, Std. Error ($S_{\hat{c}}$) = 0.3161, $t_{\hat{c}}$) = 5.153).

As well The average effect equal:

$$\hat{a}_{1\ \text{indirect effect}} * \hat{b}_{1\ \text{indirect effect}} + \hat{a}_{2\ \text{indirect effect}} * \hat{b}_{2\ \text{indirect effect}} = 1.83675 + 0.609391 = 2.446$$

$$\hat{\mu}_{\text{direct effect}} - \hat{\mu}_{\text{partial effect}} = 4.074 - 1.628 = 2.446$$

So

$$\hat{a}_{1\ \text{indirect effect}} * \hat{b}_{1\ \text{indirect effect}} + \hat{a}_{2\ \text{indirect effect}} * \hat{b}_{2\ \text{indirect effect}} = \hat{\mu}_{\text{direct effect}} - \hat{\mu}_{\text{partial effect}}$$

Standard errors , can be calculated using the standard error equation to estimate the mean effect as shown in the following:

$$S_{\hat{a}1\hat{b}1} = \sqrt{s^2_{\hat{a}1}\hat{b}^2_1 + \hat{a}^2_1 s^2_{\hat{b}1}}$$

$$= \sqrt{(0.1647)^2 (0.8646)^2 + (0.1241)^2 (2.1244)^2}$$

= 0.2996376

$$S_{\hat{a}2\hat{b}2} = \sqrt{s^2_{\hat{a}2}\hat{b}^2_2 + \hat{a}^2_2 s^2_{\hat{b}2}}$$

$$\sqrt{(0.005894)^2 (10.1430)^2 + (3.4672)^2 (0.060080)^2} =$$

= 0.2167182

$$S_{\hat{a}1\hat{b}1} + S_{\hat{a}2\hat{b}2} = \sqrt{s^2_{\hat{a}1\hat{b}1} + s^2_{\hat{a}2\hat{b}2} + 2a_1 a_2 S_{\hat{b}1\hat{b}2}}$$

$$= \sqrt{0.2996376 + 0.2167182 + 2(2.1244)(0.060080)(0.0845)}$$

= 0.5379259

The confidence limits were as follows:

Upper Confidences Limits (UCL) = Mediated effect + $Z^*S_{\hat{a}1\hat{b}1}$

= 2.4461+ 1.96 (0.2996376)

ISSN NO: 2769-996X
Date of Publication: 13-04-2022

= 3.0333

Lower Confidences Limits (LCL) = Mediated effect – Z*$S_{\hat{a}1\hat{b}1}$

= 2.4461- 1.96 (0.2996376)

= 1.8588

Upper Confidences Limits (UCL) = Mediated effect + Z*$S_{\hat{a}2\hat{b}2}$

= 2.4461+ 1.96 (0.2167182)

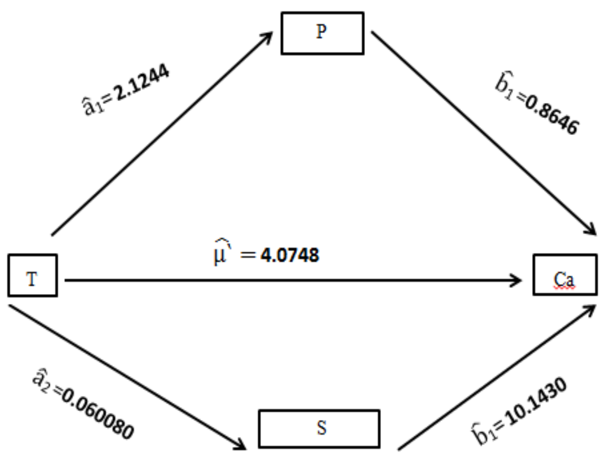
= 2.87086

Lower Confidences Limits (LCL) = Mediated effect – Z*$S_{\hat{a}2\hat{b}2}$

= 2.4461- 1.96 (0.2167182)

= 2.021332

All results show that there is statistical significance and significance of the effect of the variables (T, P, S , Ca)



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

B - Bootstrapping Estimation

To achieve greater accuracy, in the smoothing, 1500 samples were taken from 5000 cases with replacement from the original sample and each indirect effect was calculated. The appendix includes an R code with a sample Bootstrapping command.

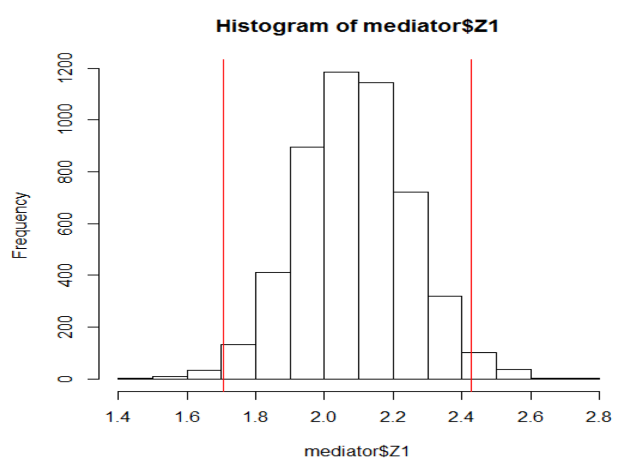
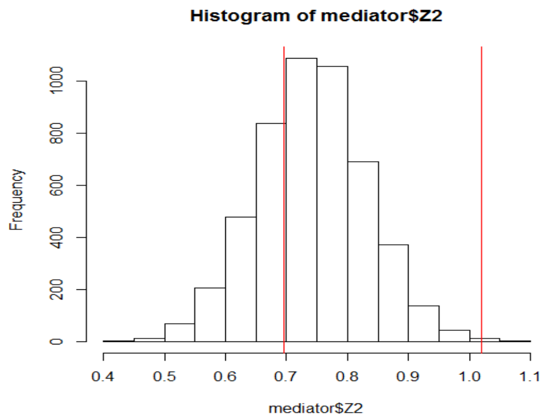


Figure (4): The diagram of the effect is shown by Z1. The red lines represent the minimum and the highest for every 95% of the confidence interval.

From the above graph we see the confidence interval for variable Z1 with a 95% confidence interval. This is because the period is limited between (1.3 - 2.4), that is, zero does not belong to the limits of this period, so it can be said that the effect of the mediation variable is different from zero and therefore there is an effect for this variable.



Figure(5): The diagram of the effect is shown by Z2. The red lines represent the minimum and the highest for every 95% of the confidence interval.

From the above graph we see the confidence interval for variable Z1 with a 95% confidence interval. This is because the period is limited between (0.2 - 1.1) ,that is, zero does not belong to the limits of this period, so it can be said that the effect of the mediation variable is different from zero and therefore there is an effect for this variable.

Conclusion

1- The effect of the cause variable on the dependent variable (total effect) was significant and (4.0748).
2- The effect of temperature on the dependent variable (partial effect) with the presence of the mediation variable significantly (4.0748).
3- The effect of the first mediation variable (population) on the dependent variable (calcium element) was significant and (0.8646).
4- The effect of the second mediating variable (soil) on the dependent variable (calcium element) was significant and (10.1430).

By observing the results in the practical side, it is clear that the Bootstrapping method is better than the product of parameter coefficients testing .

.

Reference

1. Briggs, N. E. (2006). Estimation of the standard error and confidence interval of the indirect effect in multiple mediator models. The Ohio State University.
2. Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 227–246.
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. Harlow, L. L. (n.d.). Mulaik. SA, and Steiger, JH, Eds., 1997. *What If There Were No Significance Tests*.
5. Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420.
6. MacKinnon, D. (2012). *Introduction to statistical mediation analysis*. Routledge.
7. MacKinnon, D. P. (2000). Contrasts in multiple mediator models. *Multivariate Applications in Substance Use Research: New Methods for New Questions*, 141–160.

8. 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.
9. 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.
10. Wen, S. (2013). Estimation of multiple mediator model.
11. Wu, H., Dwyer, K. M., Fan, Z., Shircore, A., Fan, J., & Dwyer, J. H. (2003). Dietary fiber and progression of atherosclerosis: the Los Angeles Atherosclerosis Study. *The American Journal of Clinical Nutrition*, *78*(6), 1085–1091.