Testing and Interpreting Correlation, Moderation and Mediation Effects in Social Science Researches

T. L. Sajeevanie

Department of Human Resource Management, University of Sri Jayewardenepura, Sri Lanka

Abstract: **Most of the researchers in positivist paradigm develop conceptual frameworks to test the research hypotheses. In a quantitative study it is very common to identify the relationship between independent and dependent variables. In addition to that, most of the conceptual models consist of moderating and mediating variables. Hence, it is very critical to test these effects and similarly it is very serious, as to how to interpret, the test results properly. Hence, the objectives of this study are; to explain the correlation, moderating and mediating effects; to explain how to interpret the test results of correlation, moderation and mediating effects in a social science research.**

Keywords: **Correlation, Moderation, Mediation, Testing, Interpretation**

I. INTRODUCTION

Most of the quantitative researchers consider the testing of relationship using correlation analysis, moderating testing and mediating testing. However, some students do not have proper awareness regarding these variables. The main purpose of this article is to explain the correlation, mediation, and mediation effects and discuss the way of interpretation of the analyzed results.

Objectives

1. To explain what is correlation.
2. To identify the moderating and mediating effects.
3. To explain how to interpret the correlation, moderation and mediation analysis results.

II. LITERATURE REVIEW

Correlation

In general correlation means a mutual relationship between two things. Correlation in the broadest sense is a measure of an association between variables. Statistically it can be explained as correlation is a term that is a measure of the strength of a linear relationship between two quantitative variables. This relationship may be of positive or negative correlations. The most common measure of correlation is Pearson's correlation. The result of this analysis is 'r', that is correlation coefficient. The most important factor is correlation coefficient 'r', which measures the strength and direction of a linear relationship.

How to interpret the correlation coefficient

'1' indicates a perfect position correlation

'-1' indicates a perfect negative correlation

'0' indicates that there is no relationship between the two variables.

Other levels of Correlation Coefficient have been interpreted by different authors differently. A common rule of thumb specially in academic research, is that when P value is greater than 0.05 the correlation should not be trusted. Other most important factor is a single unusual observation (outlier) can make the complete correlation coefficient highly misleading.

Here, positive correlation is a relationship between two variables in which both variables move in the same direction. That means when one variable increases while the other increases and vice versa. Negative correlation is a relationship where one variable increases as the other decreases and vice versa. In figure 1, it shows positive perfect correlation (r=+1), strong positive association (r close to +1), strong negative association (r close to -1), and weak or no association (r close to o).

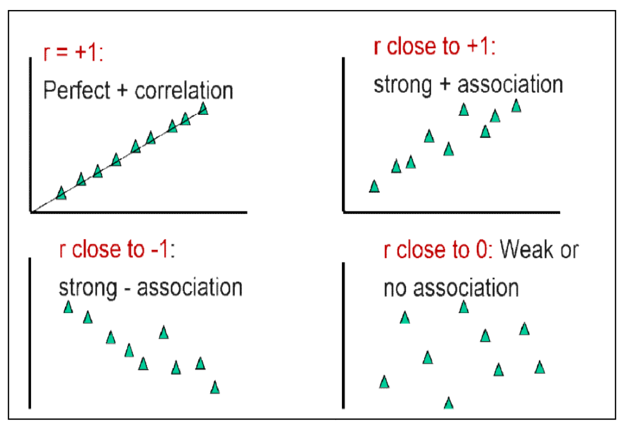

Figure 1: Positive, Negative and Zero Correlations

Moderator

As Frazier, Tix and Barron (2004, p. 116) defined, a moderator is a "variable that alters the direction or strength of the relation between a predictor and an outcome". Further, this can be explained as an interaction whereby the effect of one variable depends on the level of another.

On one hand Hair et al, (1998) explained that moderating effect occurs when a third, variable or construct changes the relationship between two related constructs. On the other hand, Lai (2013) explained that a moderator is an independent variable which affects the strength and/or direction of the connotation between another independent variable and an outcome variable. Similarly, the term moderating variable refers to a variable that can strengthen, diminish, negate, or otherwise alter the association between independent and dependent variables. Furthermore, moderating variables can change the direction of the relationshipA moderating variable may be a qualitative (religion, motivation, satisfaction) or quantitative (age, salary, price) one that affects the strength and/or direction of the relationship between dependent and independent variables (Namazi and Namazi, 2016).

This can be further explained as, the moderating variable lies in identifying whether the relationship between the predictor and criterion variable differs for a particular group or not. For example if a researcher wants to study the relationship between take-home salary and employee performance level then it is critical to identify whether this relationship is moderated by some other variables, such as gender of the employee, marital status of the employee.

Another important factor is that the moderating variable should be chosen with strong theoretical support. It is required to identify some logical reason and prior theoretical support for 'why a certain variable is likely to affect the hypothesized relationships between the variables. Kim et al, (2001) cited by Rayees and Sandeep (2017) stated that moderating variable can be at ratio, interval or continuous level or it can be categorical. Consequently this effect it is needed to be tested in order to establish the moderating effect empirically. Such interaction effects (moderators) are very important to be studied (Frazier et al., 2004). This identification of important moderators of the relations between two variables indicates the maturity and sophistication of a field of inquiry (Frazier et al., 2004).

Mediator (Intervening variable)

Mediator can be defined as a "variable that explains the relation between a predictor and an outcome (Frazier et al., 2004, p. 116). Hence, a mediating variable is a variable that links the independent and the dependent variable and this explains the relationship between the other two variables. A mediating variable is also known as a mediator variable or an intervening variable.

As Frazier et al (2004) mentioned, a researcher can examine mediators and moderators within the same model. This refers to moderated mediation. One can explain this as instances in which the mediated relation varies across levels of a moderator.

III. ANALYSIS

Figure 2 is an illustration of a format of a conceptual framework. This shows an independent variable (Practices of HRM-PHRM), a dependent variable (Organizational Performance - OP) and there is another variable that intervenes the relationship between independent and dependent variable. It's called Mediating variable (Eg: Employee Motivation- EM). There is another variable called moderating variable (Eg: Employee Competency-EC) which moderates the relationship between independent and mediating variable. This model exhibited the correct shape of the variable when presented in a conceptual model.

3.1 A Model showing the relationship of Independent, Dependent, Moderating and Mediating variables

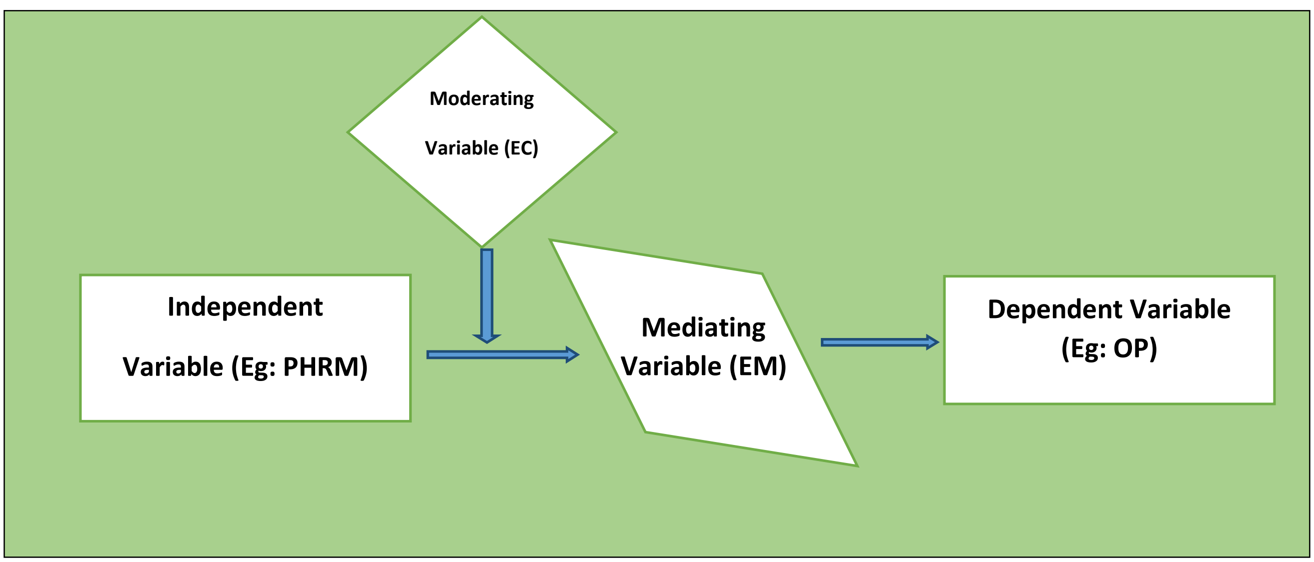

Figure 2: A conceptual model showing the relationship of Independent, Dependent, Moderating and Mediating variables.

Here, as explained above "Whether a variable is an independent variable, a dependent variable, a mediating variable , or moderating variable should be determined by a careful reading of the dynamics operating in any given situation. For instance, a variable such as motivation to work could be a dependent variable, an independent variable, a mediating variable, or a moderating variable, depending on the theoretical model that is being advanced" (Sekaran and Bougie, 2013, p76).

3.2 How to interpret tested results of a conceptual model

The major part of quantitative research study is to test the conceptual model with hypotheses testing. First the researcher can use Pearson Correlation analysis to test the linear relationship between the constructs, and then used linear multiple regression analysis. In regression analysis, statistical tests such as R2 (coefficient of multiple determination) individual regression coefficients of independent variables, β (beta) and F statistics are used. Accordingly, the significance of regression coefficients is tested at the $\alpha = 0.05$ comparing the "P" probability value generated by SPSS outputs. Moreover, it is needed to test the mediating and moderating effect of the variables. Multiple regression analysis technique can be used following the procedure recommended by Baron and Kenny (1986)

Multivariate Assumptions

In order to test the appropriateness of the data for analysis first the following assumptions need to be tested.

(1) Normality

Normality is that the distribution of the test is normally distributed with 0 mean, with 1 standard deviation and a symmetric bell shaped curve. Basically to test the assumption of normality the following measures and tests are applied;

(a) Skewness- to test the assumption of normal distribution, skewness should be within the range ± 1.

(b) Kurtosis – to test the assumption of whether or not the data is normally distributed, kurtosis value should be within the range of ±3. Some authors have identified it as ±2 range of kurtosis.

Further, it can be explained as normality magnifies the shape of the sample data distribution to the population. Data need to be of approximately normal distribution. According to Hair et al., (1998) data distribution either with a highly skewed nature or with high kurtosis is indicative of non-normality, which has random effects on specification or estimation. As they further explained this non-normality may exist due to the presence of outlier cases in the data set.

(2) Linearity

Linearity is the amount of deviation from an instrument's ideal straight-line performance. It investigates the presence of a straight-line relationship between two variables. This relationship can be expressed as an equation:

$$y = a + \beta_1 x_1 + \cdots + \beta_n x_n + \acute{\varepsilon}$$

(Sekaran at el., 2010)

(3) Homoscedasticity

Homoscedasticity is also referred to as homogeneity of variance or uniformity of variance. As linearity, there are two methods for evaluating homoscedasticity called graphical and statistical methods. The graphical method is called a boxplot and scatterplot. The statistical method is the Levenes Statistic which SPSS computes the test of homogeneity of variance

(4) Multicollinerity

Nonexistence of multicollinerity is another important assumption to be met. "Multycollinearity is an often encountered statistical phenomenon in which two or more independent variables in a multiple regression model are highly correlated" (Sekaran at el, 2010, p355). This can be further explained as high intercorrelations among the independent variables making it very difficult to ascertain and separate the influence of a single independent variable on a dependent variable. Further Sekaran at el (2010) explained that the simplest and most obvious way to detect multicollinrarity is to check the correlation matrix for the independent variables. Basically, the correlation matrix of independent variables is useful to determine inter-correlations ($p<.09$). However, as Sekaran et al (2010) explained when multicollinearity is the result of complex relationships among several independent variables, it may not be revealed by this approach. According to them more common measures for identifying multicollinearity are the tolerance value and the variance inflation factor (VIF). These measures indicate the degree to which one independent variable is explained by the other independent variables. The cutoff value is a tolerance value of 0.10, which corresponds to a VIF of 10. As explained by Sekaran et al (2010) if the tolerance value is less than 0.10 then the multicollinearity problem should be taken into consideration and VIF more than 10 also indicates that there is a multicollinearity issue.

3.3. How to interpret the Correlation analysis

"A Pearson correlation matrix will indicate the direction, strength, and significance of the bivariate relationship among all the variables that were measured at an interval or ratio level" (Sekaran at el, 2010, p289). Similarly, as Sekaran very clearly explained "A correlation coefficient that indicates the strength and direction of the relationship can be computed by applying a formula that takes into consideration the two sets of figures".

Following example, explains the way of interpreting correlation analysis.

e.g. Pearson correlation was used to test the null hypothesis. One tailed test was used because the bivariate hypothesis is concerned with a positive relationship. The Pearson correlation matrix of the variables investigated is shown in table 1 and 2. According to the table 1 the practice of HRM and Employee Motivation has positively correlated with a coefficient of .446 and the relationship is significant at the 0.01 level. As table 2 depicts organizational performance and

Practice of HRM have positively correlated with a coefficient of .373 and the relationship is significant at the 0.01 level. Hence, the null hypotheses proposed can be rejected. Accordingly, it can be concluded that hypothesized relationship (positive) between each variable was supported by the data.

Table 1: The Relationship Between Practice Of Hrm And Employee Motivation

		PHRM	EM
PHRM	Pearson Correlation	1	..446**
	Sig. (2-tailed)		.000
	N	301	301
**. Correlation is significant at the 0.01 level (2-tailed).			

Table 2: The Relationship Between Practice Of Hrm And Organizational Performance

		PHRM	OP
PHRM	Pearson Correlation	1	.373**
	Sig. (2-tailed)		.000
	N	301	301

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation Coefficient

A correlation coefficient indicates the strength and direction of a relationship between two variables. In order to statistically accept the relationship, significant level should be considered. A significance of P=0.05 is the generally accepted conventional level in social science research. This has been explained by Sekaran and Bougie (2013, p. 290) as “95 confidence level mean 95 times out of 100, we can be sure that there is a true or significant correlation between the two variables, and there is only a 5% chance that the relationship does not truly exist”.

Correlation coefficient can be interpreted in a different way. Different authors have described different threshold levels. The following table 3 shows Schcher et. al. (2018) way of interpreting correlation coefficient.

Table 3: An Interpretation of Correlation Coefficients

Correlation Coefficients	Interpretation
0 - .10	Negligible Correction
0.10 – 0.39	Week Correlation
0.40 – 0.69	Moderate Correlation
0.70 – 0. 89	Strong Correlation
0. 90 – 1.00	Very Strong Correlation

Source: Schcher et al. (2018). Correlation Coefficient: Appropriate use and interpretation

3.4 How to interpret the Mediating effect

Following example, explains the way of interpreting mediating according to Frazier et al. (2004).

e.g. Mediating Effect of Employee Motivation on the Relationship Between Practice of Human Resource Management and Organizational Performance

The hypothesis of this research is ‘Employee Motivation (EM) has a significant mediator effect on the relationship between Practice HRM (PHRM) and Organizational Performance (PO)’.

According to Frazier et al. (2004) there are four steps in establishing that a variable mediates the relationship between a predictor variable and an outcome variable.

1. The first step is to show that there is a significant relationship between the predictor (Practice of HRM -PHRM) and the outcome (Organizational Performance-OP).

Table 4: Predictor (Phrm) Outcome(Op) Relationship –Regression 1

Construct	B	R squire	Sig
PHRM and OP	.465	.415	.000

2. The second step is to show that the predictor (Practice of HRM) is related to the mediator (Employee Motivation- EM)

Table 5: Predictor And Mediator Relationship (Phrm And Em) – Regression 2

Construct	B	R squire	Sig
PHRM and EM	.566	.425	.000

3. The third step is to show that the mediator (e.g. Employee Motivation) is related to the outcome variable (Organizational Performance)

Table 6: Mediator And Outcome Relationship (Em & Op) – Regression 3

Construct	B	R squire	Sig
EM and OP	.208	.367	.000

4. The final step is to show that the strength of the relationship between the predictor (Practice of HRM) and the outcome (Organizational Performance) is significantly reduced when the mediator (Employee Motivation) is added to the model.

Table 7: The Strength Of The Relationship Between The Predictor And The Outcome – Regression 4

Construct	B	R squire	Sig
PHRM and OP When EM is added to the model	.325	.412	.000

As mentioned above the first step is to show that there is a significant relationship between predictor (Practice of HRM) and the outcome (Organizational Performance). According to regression 1 (Table 4) there is a significant relationship between PHRM and OP (B=.465, p<0.000). The second step is to show that the predictor (Practice of HRM) is related to the mediator (Employee Motivation). In regression 2 (Table 5) there is a significant relationship between Practice of HRM and Employee Motivation (B=.534, p>0.000). Third step is to show that the mediator (Employee Motivation) is related to the outcome (Organizational Performance) variable. In regression 3 (Table 6) there is a significant relationship between Employee Motivation and Organizational Performance (B=0.208, p<0.000). The final step is to show that the strength of the relationship between the predictor (Practice of HRM) and the outcome (Organizational Performance) is significantly reduced when the mediator (Employee Motivation) is added to the model. In regression 4 (Table 7)(the overall model is significant, p value of F statistic<0.05 and R2>0), the unstandardized regression coefficient (B) of Employee Motivation is also significant (B= 0.272, P<0.05). Further, when comparing with B (0.465) of regression 1 the B of PHRM in regression 4 has reduced to 0.325 which is significant. Thus, all the conditions necessary for a mediated relationship are satisfied. Since the B value of practice of HRM in regression 4 is still significant, based these it can be concluded that Employee Motivation has a mediating effect on the relationship between Practice of HRM and OP. However, it is not acting as a perfect mediator. Further, As Frazier et al. (2004) mentioned the study uses the raw regression coefficient (B) for the mediator testing.

3.5 How to interpret the Moderating effect

Following example, explains the way of interpreting the moderating effect according to Frazier et al. (2004).

Moderating Effect of Employee Competency (EC) on the Relationship between Practice of Human Resource Management and the Employee Motivation

Second Hypothesis of this study is: Employee Competency (EC) has a significant moderating effect on the relationship between Practice of HRM and the Employee Motivation (EM).

The moderation model tests whether the prediction of a dependent variable Y from an independent variable, X differs across levels of a third variable, Z (Fairchild & Mackinnon, 2009). As they explained moderator variable affects the strength and/or direction of the relation between a predictor and an outcome: enhancing, reducing, or changing the influence of the predictor. As mentioned, presented in the figure number two, there is a moderating variable in the research model. Employee Competency is the moderating variables which moderate the relationship between Practice of Human Resource Management and Employee Motivation. In order to analyze the moderating effect Baron et al. (1986) mentioned the need to test the impact of predictor, moderator, and the interaction or product of these two on the outcome variable. They have clearly mentioned that the moderator hypothesis is supported if the interaction is significant. Further, they have mentioned that there may also be significant main effects for the predictor and the moderator but these are not directly relevant conceptually to testing the moderator hypothesis (Baron et al., 1986, p. 1174).

According to the Hypothesis, Employee Competency moderates the relationship between Practice of HRM and the Employee Motivation. As explained by Baron et al. (1986), where both independent and dependent variables are continuous, the moderating hypothesis is tested by adding the product of the moderator and the independent variable to the regression equation. Hence, if the independent variable is denoted as X, Y is the dependent variable, the moderator is Z and XZ acts as the product of X and the moderator variable. This is a dummy variable, representing the moderator is multiplied by the independent variable. When analyzing moderation effect Fairchild et al. (2009) expressed a single regression equation to form the basic moderation model: Y = i5+β1X+ β2Z+ β3XZ+e5. Here, β1 is the coefficient relating to the independent variable, X to the outcome Y, when Z=0 (the effect of X on Y), β2 is the coefficient relating the moderator variable Z to the outcome when X=0 (the effect of Z on Y), i5 intercept in the equation, and e5 is the residual in the equation. Further, they have guided as the regression coefficient for the interaction term, β3 to provide an estimate of the moderation effect. If β3 is statistically different from zero, there is significant moderation of the X-Y relation in the data. In order to test the moderating effect, first computes the moderating effect variable (EC) by multiplying independent (Practice of HRM) variable with moderator (Employee Competency). (When computing moderating variable the data set was divided into two as; Group 1 High >3.7 and Group 2 Low< 3.7 created a new dummy variable. This new dummy variable representing the moderator is multiplied by independent variable (PHRM *ECM Dummy) and resulting interaction variable (PHRMEC) was included in the regression analysis. The results were as follows:

Table 8: Moderating Effect of Employee Competency

	Beta	Sig
PHRM	.560	0.003
PHRMEC	.768	0.001
ECM	-	Not Significant

Overall model of F value was significant and R =.63 and r2 =.396. Moreover, the main effect of Practice of HRM on Employee Motivation is statistically significant (Please refer Appendix) and the interaction variable (PHRMEC) is significant at 99% level which is highly aligned with Baron et al. (1986, P1174) as they mentioned "The moderator hypothesis is supported if the interaction (path C) is significant." Thus there is statistical evidence to conclude that the relationship between Practice of HRM and Employee Motivation is moderated by Employee Competency.

IV. CONCLUSION

As explained above correlation, moderation and mediating effect are important for social science quantitative researchers. However some of them do not have very clear idea of correlation, moderating, mediating effects. Similarly some researchers don't clearly interpret the analyzed data relating to correlation analysis, moderating analysis and mediating analysis. Hence, it is strongly suggesting to carefully interpreting the analyzed data. This way of interpreting the analyzed data can be used for quantitative research students of undergraduates, Master Degree students and any other scholars for their studies.

REFERENCES

[1] Fairchild A.J. and Mackinnon, D. P. (2009). A General Model for Testing Mediation and Moderation Effects. *Prevention Science*, Vol. 10, 87-89.

[2] Frazier. P.A., Brron, K.E. and Tix, A.P (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research, *Journal of Counseling Psychology,* Vol.51(1), 115-134.

[3] Hair, J.F.J, Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2006). *Multivariate data analysis* (6^{th} ed.). New Jersey: Prentice Hall.

[4] Namazi, M. and Mamazi, N. (2016). Conceptual Analysis of Moderator and Mediator Variables in business research. *Procedia Economics and Finance,* Col (36), 540-554.

[5] Rayees, F. and Sandeep V., (2017). Moderating Variables in Business Research. *IUP Journal of Business Strategy,* 14(4), 75-79.

[6] Schcher, P., and Schwarte L.A. (2018). Correlation Coefficients: Appropriate use and Interpretation . *Anesthesia and Analgesia,* 126(5), 1763-1768.

[7] Sekaran, U. and Bougie, R. (2013). Research Methods for Business; A skill-Building Approach. Wiley India Pvt.Ltd, New Delhi

Annexe

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.630	.396	.397	.48820

a. Predictors: (Constant), PHRM, Employee Competency, PHRMEC

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.012	.325		9.270	.000
	PHRMEC	.757	.017	.768	3.323	.001
	PHRM	.510	.070	.560	3.004	.003

a. Dependent Variable: EM

ANOVA[b]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13.068	3	4.356	48.276	.000[a]
	Residual	70.786	297	.238		
	Total	83.854	300			

a. Predictors: (Constant), PHRM, , PHRMEC

b. Dependent Variable: EM