Interpreting the Basic Results of Multiple Linear Regression

Ganesh Prasad Adhikari

Central Department of Mathematics Education

Tribhuvan University, Kirtipur, Nepal

gpadhikarin@gmail.com

Abstract

Regression analysis is one of the most useful tools for academics, although it is a difficult, time-consuming, and expensive effort, especially when it comes to accurately estimating and properly interpreting data. Researchers believe that the findings of multiple linear regression produced by SPSS require a more inclusive and thoughtful interpretation. This study aims to understand and illustrate the detailed interpretation of fundamental multiple linear regression results using the social science sector. In this paper, researcher describe the processes for using SPSS Version 26 to obtain the results from multiple linear regression, and we also show the detailed interpretation of the results. In the results, Model Summary table, Statistical Significance of the Model from the ANOVA Table, and Statistical Significance of the Independent Variables from the Coefficients Table, researcher illustrate the interpretation of the coefficient from the output, B-value, β-value, t-value, and p-value. The results of multiple regression have been discussed in a thorough and careful manner. The resultant multiple linear regression model's statistical and substantive significance are discussed. Every effort has been made to ensure that the explanation of the findings throughout the study serves as a competent model for the researchers to apply to any real-world data they may employ. Any researcher using multiple linear regression to more accurately predict their outcome variable should feel at peace and gain from doing so because every effort has been made to properly comprehend the fundamental SPSS multiple linear regression outputs.

Keywords: Normality, independence, outliers, homoscedasticity, and multicollinearity.

Introduction

A multiple linear regression analysis is a statistical method for creating a model and examining the relationship between the dependent and independent variables (Chatterjee & Hadi, 2012a). The dependent variable is referred to as an explanation, response, predicted, or regress and variable, while the independent variable is also known as an explanatory, control, predictor, or regressor variable (Chatterjee & Hadi, 2012a; Cohen, 1988a). The goal of multiple linear regression is to ascertain the degree of relationship between two or more variables and it involves predicting values for a dependent variable (Y) given a collection of kth predictor variables ($X_1, X_2, X_3, ..., X_k$). Additionally, we also use it when we want to determine which variables are better predictors than others. The relationship between the dependent variable and the explanatory variables in multiple linear regression is represented by the following equation, where there are p explanatory variables: Y= β_0 + $\beta_1 X_1$+ $\beta_2 X_2$ + $\beta_3 X_3$ +…+ $\beta_p X_p$ + e, where β_0 is the constant term and β_1 to β_p be the coefficients relating the *p* explanatory variables to the variables of interest. Therefore, multiple linear regression may be thought of as an extension of simple linear regression when there are p explanatory variables, or simple linear regression can be thought of as a specific case of multiple linear regression when p=1. The word "linear" is employed because, in multiple linear regression, we presume that Y is connected to a linear combination of the explanatory variables. Multiple linear regression is used in various social science research, but it is not used correctly (Seawright, 2005; Sinharay et al., 2001; Uyanık & Güler, 2013). That's why in this article, the researcher gives the appropriate guidelines and discusses prerequisite assumptions that must be satisfied before performing multiple linear regression and the interpretation criteria for using them correctly by using SPSS software version 26.

Assumptions

- Continuity: the dependent and independent variables should be measured on a continuous scale.
- Linearity: there need to be linear relationship between the dependent variable and independent variables.
- Normality: the dependent variable and independent variables should be normal. i. e. residuals/error terms should be normally distributed with a mean of zero and constant variance e ~N(0, σ^2).
- Independence: all of the observations are independent.
- Outlier: there should be no significance outliers.
- Homoscedasticity: the variance of residual terms should be the same at each level of the predictor variable(s).
- Multicollinearity.

The findings of our multiple linear regression might be inaccurate or even deceptive if one or more of these underlying assumptions is violated. Therefore, in this article, researcher explain each assumption and discuss how to check whether it is true, as well as what to do if it is not, by taking one example using SPSS software version 26.

Methods and Procedures

This article is based on books, journal publications (original research articles, and review articles). These were selected from Google Scholar, Science Direct, and Library Genesis databases, which were used to search the literature. Additionally, the standardized questionnaires were administered to 126 randomly selected secondary mathematics teachers in the Kathmandu district. The standardized questionnaire comprises seven dimensions as shown in the **Figure 1** for gathering the necessary data to apply multiple linear regression. Therefore, the researcher used the SPSS-26 version software to analyze the dataset obtained from 126 teachers, checking it for linearity, normality, independence, outliers, homoscedasticity, and multicollinearity by correlation, Skewness and Kurtosis, Durbin-Watson test, boxplot, Mahalanobis Distance Test, scatterplot, and tolerance and VIF tests, respectively.

Figure 1

Regression Model
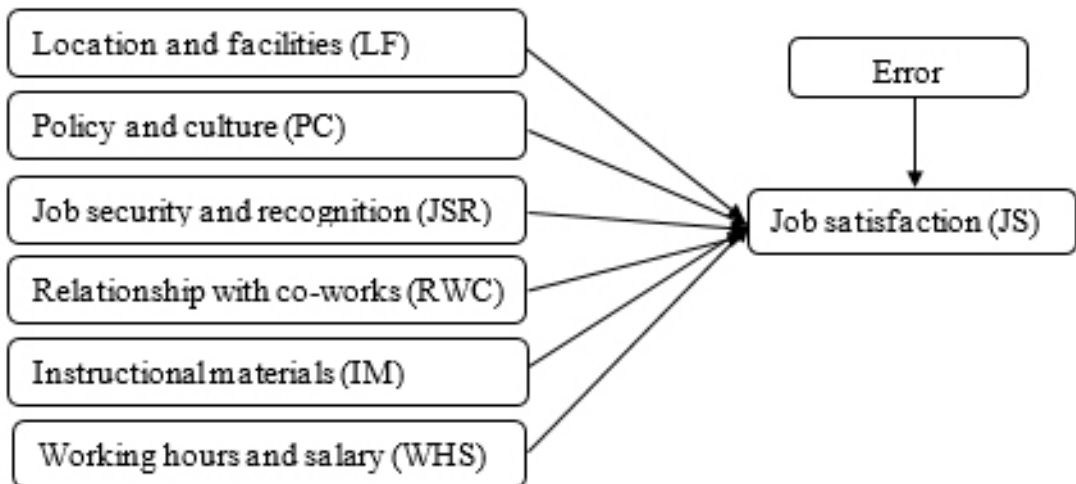
Thus, the required regression equation (model) of this study is

Job satisfaction (Y) $= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + + \beta_5 X_5 + \beta_6 X_6 + e$, where

$X_1 =$ Location and facilities (LF),

X_2 = Policy and culture (PC),

X_3= Job security and recognition (JSR),

X_4 = Relationship with coworkers (RWC),

X_5= Instructional materials (IM), and

X_6= Working hours and salary (WHS).

The researcher used the following research hypothesis to discuss the model's hypotheses for this study:

H_0: There is no statistically significant impact of the independent variables (LF, PC, JSR, RWC, and WHS) on dependent variable (JS) of secondary mathematics teachers in Kathmandu district.

H_a: There is statistically significant impact of the independent variables (LF, PC, JSR, RWC, and WHS) on dependent variable (JS) of secondary mathematics teachers in Kathmandu district.

Results and Discussion

The following pre-assumptions underlie multiple linear regression:

Continuity: Since all the dependent and independent variables are continuous in the given data because the researcher calculated the mean scores from the five-point Likert scale (Chatterjee & Hadi, 2012b; Shapiro et al.,1968). Therefore, the first assumption is to fulfill the requirement to apply multiple linear regression.

A linear relationship between the dependent and independent variables: This assumption is the most important because if it is not true then even if all other assumptions are meet, your model is invalid because you have described it incorrectly (Chatterjee & Hadi, 2012b; Field, 2009). Thus, dependent variable and each of the independent variables must have a linear relationship in order for multiple linear regression to do. There are so many methods for checking the linearity, which are scatter plot, correlation, histogram, normal Q-Q plot, and box plots. In this study, the data was checked by the correlation method, which is shown below:

Table1

Correlations between the dependent variable and independent variables

		JS	LF	PC	JSR	RWC	IM	WHS
	JS	1.000						
	LF	.489	1.000					
	PC	.660	.424	1.000				
Pearson Correlation	JSR	.582	.371	.485	1.000			
	RWC	.535	.350	.681	.365	1.000		
	IM	.561	.307	.708	.467	.474	1.000	
	WHS	.348	.295	.409	.525	.280	.312	1.000

Since, from the Table 1 shown that the correlation lies within the range of 0.3 to 0.7. Thus, the linear relationship between the dependent variable and the independent variables holds. If the correlation does not lie within the range of 0.3 to 0.7 (Meng et al., 1992), then the linearity does not hold for the dependent variable and independent variables.

Normality: The assumption of normality plays a vital role in data analysis (Chatterjee & Hadi, 2012b). We go for a parametric test if the data is normal and a non-parametric test if the data is non-normal. For the multiple linear regression, the data should be normal. If the date is not normal, we cannot use it. Therefore, the researcher should check the normality of the data. There are various methods to check if the data is normal or not, which are the graphical method, the Shapiro-Wilk test, the Kolmogorov Smirnov test, the Skewness test, and the Kurtosis test. For testing the normality of data, it is best to test by statistical method (Shapiro et al., 1968).

Skewness and Kurtosis: If the sample size is less than 50, then the Skewness z-value and Kurtosis z-value lie within the range of -1.96 to 1.96 (the Skewness z-value and Kurtosis z-value are found by the Skewness and Kurtosis measure divided by its standard error). If the sample size is between 50 and 300, then we use a more liberal z-range of -3.29 and 3.29. If the sample is greater than 300, then the absolute skewness value will be between -2 and 2, and the absolute kurtosis value between -7 and 7 (Shapiro et al., 1968).

Kolmogorov-Smirnov and Shapiro-Wilk's tests: These two tests will be used if the sample size is below 300. If the sample size is greater than 300, it may be unreliable. Therefore, the researcher will use the Kolmogorov-Smirnov test if the sample size lies within the range of 100 and 300. Also, Shapiro-Wilk's test is used when the sample size is less than 100 (Shapiro et al., 1968).

In this study, the data was checked by the Skewness and Kurtosis z-values tests method, which is shown below:

Table 2

The values of the Skewness/Kurtosis z-values of dependent variable and independent variables

Variables	Statistical Measure	Std. Error	Skewness/Kurtosis z-values
JS	Skewness	0.571	0.216	2.64
	Kurtosis	0.586	0.428	1.37
LF	Skewness	-0.345	0.216	-1.60
	Kurtosis	0.497	0.428	1.16
PC	Skewness	-0.593	0.216	2.75
	Kurtosis	0.467	0.428	1.09
JSR	Skewness	0.574	0.216	2.66
	Kurtosis	0.643	0.428	1.50
RWC	Skewness	0.436	0.216	2.02
	Kurtosis	0.542	0.428	1.26
IM	Skewness	0.461	0.216	2.13
	Kurtosis	0.379	0.428	0.89
WHS	Skewness	0.416	0.216	1.92
	Kurtosis	0.612	0.428	1.43

The Skewness and Kurtosis z-values revealed that the data of the dependent variable (JS) and independent variables (LF, PC, JSR, RWC, IM, WHS) were approximately normally distributed, with Skewness of 0.571 (SE = 0.216) and Skewness z-value 2.64, and Kurtosis of 0.586 (SE=0.428) and Kurtosis z-value 1.37 for the dependent variable (JS). Similarly, Skewness of -0.345 (SE = 0.216) and Skewness z-value -1.60, and Kurtosis of 0.497 (SE = 0.428) and Kurtosis z-value 1.16 for the independent variable LF. Additionally, for the independent variable (PC), Skewness was -0.593 (SE = 0.216) and the Skewness z-value was 2.75. Kurtosis was 0.467 (SE = 0.428) and the Kurtosis z-value was 1.09. Also, Skewness of 0.574 (SE=0.216) and Skewness z-value 2.66, and Kurtosis of 0.643 (SE=0.428) and Kurtosis z-value 1.50 for the independent variable JSR. Moreover, Skewness of 0.436 (SE=0.216) and Skewness z-value 2.02, and Kurtosis of 0.542 (SE=0.428) and Kurtosis z-value 1.26 for the independent variable RWC. Skewness of 0.461 (SE=0.216) and Skewness z-value 2.13, and Kurtosis of 0.379 (SE=0.428) and Kurtosis z-value 0.89 for the independent variable RWC. Skewness of 0.416 (SE=0.216) and Skewness z-value 1.92, and Kurtosis of 0.612 (SE=0.428) and Kurtosis z-value 1.43 for the independent variable WHS (Cramer, 1998; Razali & Wah, 2011).Thus, from Table 2, it is shown that the Skewness z-value and Kurtosis z-value lie within the range of -3.29 to 3.29. Therefore, the given data for the dependent variable and independent variables are normally distributed in terms of Skewness and Kurtosis.

Independence: to fit the model in multiple linear regression the observations should be independent of one another (i.e. uncorrelated). In other words, the model implies that the residual values are independent. We can apply the Durbin-Watson statistic to verify this supposition. The test will provide values between 0 and 4, where a values between 0 and 2 indicating positive autocorrelation and values between 2 and 4 indicating negative autocorrelation (Chatterjee & Hadi, 2012b). The mid-point, i.e. a value of 2 indicates that there is no autocorrelation. A result that is nearer to 0 indicates a stronger positive autocorrelation, whereas a result that is nearer to 4 indicates a stronger negative autocorrelation (Savin & White, 1977). Therefore, testing the autocorrelation is required before using multiple linear regression. We cannot fit the good model in multiple linear regression if it violates this premise if your data exhibit autocorrelation. In conclusion, there are three problems if the data has autocorrelation, which are:

- Ordinary least square estimator will be inefficient, there is no longer best linear unbiased estimator.

- The variance of the regression coefficient will be biased and inconsistent.

- When you reach the autocorrelation, hypothesis testing is no longer valid.

Table 3
Model Summary[b]

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | | Durbin-Watson |
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.756[a]	.571	.550	.37458	.571	26.452	6	119	.000	1.980

a. Predictors: (Constant), WHS, RWC, LF, IM, JSR, PC

b. Dependent Variable: JS

The study of the data in Table 3 shows that there is no autocorrelation, as indicated by the Durbin-Watson value of 1.98, which is near to 2. As a result, the observations are independent to one another. As a result, using the provided data, multiple linear regression may be used to forecast job satisfaction using independent variables.

Outlier: An observation that differs abnormally from other values in a population sample taken at random is known as an outlier (Barnett & Lewis, 1994). There are various method to find the outlier by using SPSS which are given bellows:

(a) **In the data sheet at SPSS, go to analyze, regression, linear, then we get the dialog box, after that chose statistic at that box, tick the case wises diagnostic, continue. In this case, if the data has outlier, then the case wises diagnostic will come like Table**

4, otherwise not see that table in the result in SPSS. Two items (i.e. case numbers 24 and 57) in Table 4 are outliers in the given dataset.

Table 4

Case Wise Diagnostics

Case Number	Std. Residual	Job Satisfaction	Predicted value	Residual
24	4.327	2	0.37	1.63
57	4.327	2	0.37	1.63

a. Dependent Variable: JS

(b) Again, by other method the researcher can find outlier as: go to analyze, regression, linear, then we get the dialog box, after that chose save option, tick the Mahalonobis, and continue, ok at that box. At that time researcher get the new variable MAH_1(Mahalonobis) in the SPSS dataset. If the case with Mahalonobis distance greater than critical Chi-square value (see Table 5) are considered as outliers under the degree of freedom (df)= number of predictors, otherwise not.

In SPSS dataset, go to graph, legacy dialogs, boxplot, tick the summaries of separate variable, define, takes all the DV and IV variables in boxes represent, then the researcher get the Figure 2. There should not be any significant outliers in the dependent variable and independent variables if multiple linear regression is to be used to estimate the value of a dependent variable (Y) based on independent variables (X_i). The outlier depicted in Figure 2 does not exist in the study's data for the dependent or independent variables. As a result, employing multiple linear regression with the given study data does not contradict the assumption of an outlier (Barnett & Lewis, 1994; Houglin & Lglewicz, 1987; Savin & White, 1977)

Table 5
*Table of Critical Chi-Square Value*

| Df	P= 0.05	P=0.01	P=0.001	df	P= 0.05	P=0.01	P=0.001
1	3.84	6.64	10.83	8	15.51	20.09	26.13
2	5.99	9.21	13.82	9	16.92	21.67	27.83
3	7.82	11.35	16.27	10	18.31	23.21	29.59
4	9.49	13.28	18.47	11	19.68	24.73	31.26
5	11.07	15.09	20.32	12	21.03	26.22	32.91
6	12.59	16.81	22.46	13	22.36	27.69	34.53
7	14.07	18.48	24.32	14	23.69	29.14	36.12

Source: (Gupta, 2011)

Figure 2

Boxplot of the Dependent Variable and Independent variables
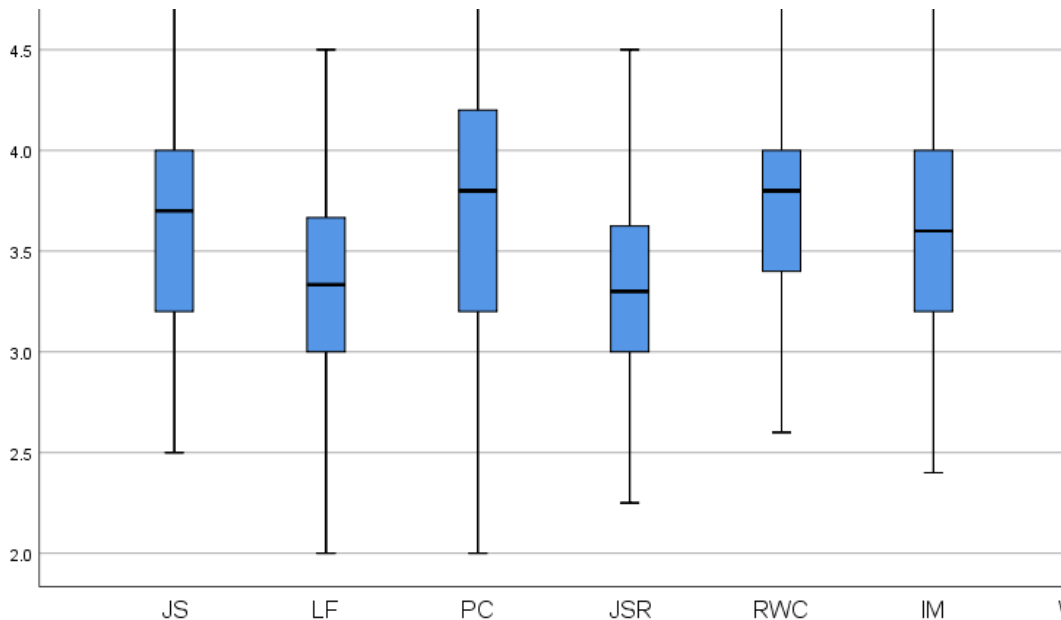
Homoscedasticity: It is an assumption in regression analysis. To understand homoscedasticity in multiple linear regression analysis, we must first understand the residual value of the dependent variable. Residual values are simple error terms. It is the difference between the observed value and the predicted value. Homoscedasticity refers to whether these residuals are equally distributed or whether they tend to cluster together at the same values and spread out at other values. If these residuals are equally distributed, it is called homoscedasticity. If the residual tends to cluster together at the same values and spread far at some other values, it is called heteroscedasticity (Breusch & Pagan, 1979). We always want to see whether this multiple linear regression analysis in the dependent variable has homoscedasticity or not. Let's use SPSS software to examine the homoscedasticity of the study's data. The steps are: analyze, regression, and linear. After then, a dialog box for linear regression appeared. Next, drag the DV into the dependent box and the IVs into the independent box, and choose plots. You may acquire the result displayed in Figure 3 by dragging the ZPRED in the X-axis and the ZRESID in the Y-axis once more.

Figure 3

Scatterplot of Regression Standardized Predicted Value and Regression Standardized Residual

Scatterplot
Dependent Variable: JS
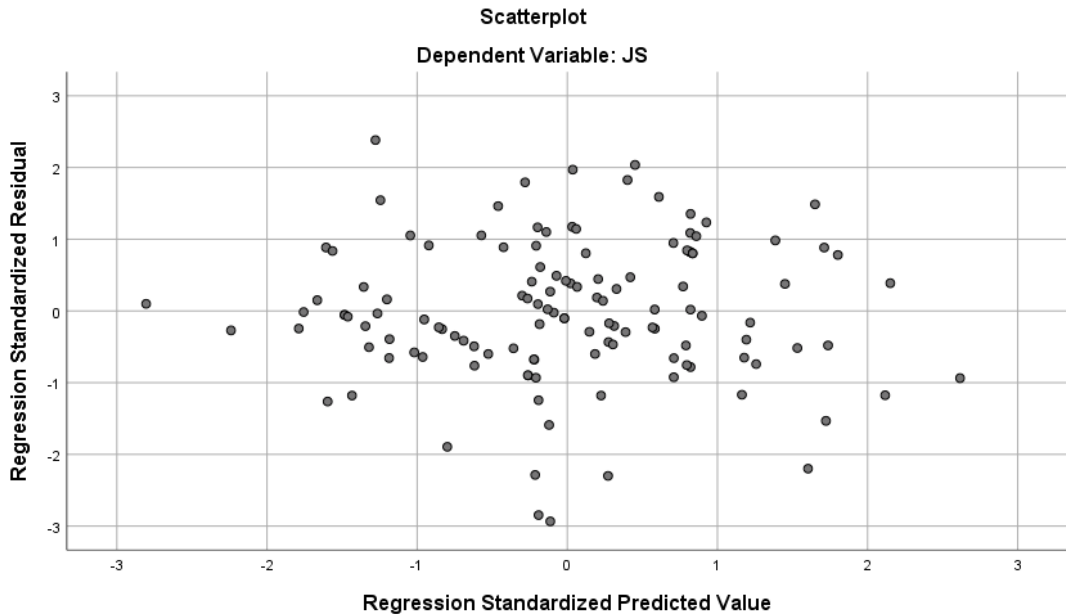
Thus, the above Figure 3 shows the distribution of the residual values. They are distributed uniformly, and we do not have any clusters forming together, so that the study data for this study is in homoscedasticity condition. As a result, the data does not violate the assumption of homoscedasticity. Hence, we can use the multiple linear regression analysis to predict job satisfaction based on the independent variables.

Multicollinearity: The term "multicollinearity" in multiple regression analysis refers to linear relationships between the independent variables. Collinearity is the relationship between two variables that is almost perfectly linear (Belsley, 1991). When a regression model has a number of variables that are substantially correlated not only with the dependent variable but also with one another, this is known as multicollinearity (Young, 2017). In any social research that employs multiple linear regression and violates this assumption will yield incorrect results, and the researcher will not fit the good model (Raykov & Marcouldes, 2006). As a result, researchers should confirm this assumption before conducting multiple linear regression. There are various methods to check the multicollinearity, which are (a) tolerance (b) variance inflation factor (VIF) (c) condition index (d) correlation analysis. Tolerance is an indicator of multicollinearity If the value of tolerance is greater than 0.2, then there is no multicollinearity problem between the independent variables. Again, the VIF is

an indicator of multicollinearity if the value of VIF is equal to or less than 10, then there is no multicollinearity; if the value of VIF is greater than 10, then there is a multicollinearity problem in the data. Similarly, if the condition index value is less than 15, then there is no multicollinearity problem. If greater or equal to 15, there is a problem. If it is greater than 30, then there is a serious multicollinearity problem in the data, and the general rule is that if the correlation between the independent variables is between -0.7 and 0.7, then there is no multicollinearity problem in the dataset (Raykov & Marcouldes, 2006). The correlation between the DV and IV should be greater than 0.3 and less than 0.7 is better. There is no issue with multicollinearity even if only one of the four conditions is met. In this case, the researcher can use multiple linear regression; otherwise, no. (Raykov & Marcouldes, 2006). There is no issue with multicollinearity even if only one of the above four conditions is met. In this case, the researcher can use multiple linear regression; otherwise, no. Now, Table 6 demonstrates that there is no multicollinearity issue since the tolerance and VIF values of the dependent and independent variables in the dataset of this study do not deviate from the assumption. Let's use SPSS software to examine the multicollinearity of the study's data. The steps are: analyze, regression, and linear. After that, a dialog box for linear regression appeared. Next, drag the DV into the dependent box and the IVs into the independent box, and choose statistic, after that we got linear regression statistic box. You may acquire the result displayed in Table 6 by clicking estimates, model fit and collinearity.

Table 6

The Tolerance and VIF Values of the questionnaire dataset

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.173	.358		-.483	.630		
	LF	.203	.072	.193	2.832	.005	.775	1.291
	PC	.262	.096	.291	2.726	.007	.315	3.175
	JSR	.411	.107	.300	3.840	.000	.592	1.691
	RWC	.149	.102	.120	1.461	.147	.531	1.884
	IM	.117	.088	.116	1.336	.184	.477	2.098
	WHS	-.055	.072	-.055	-.763	.447	.687	1.455

a. Dependent Variable: JS

As a result, the dataset for the study of secondary level mathematics teachers' job satisfaction based on the independent variables (LF, PC, JSR, RWC, and WHS) in Kathmandu district met all the requirements for applying multiple linear regression, as a

result, when conduction multiple linear regression analysis with SPSS software version 26, the researcher should adhere to the procedures specified below. First open the dataset in SPSS, after that, go to analysis, regression, linear, then linear regression dialog box comes in which the dependent variable is brought to the dependent box and the independent variables are brought to the independent box. After that, click on statistic in that box. Then another dialog box will appear in which you can click on Model Fit, Estimate, R-Square, Descriptive, and Confidence Interval. After that, the necessary results (i.e. Table 7: Model summary, Table 8: ANOVA, and Table 9: Coefficient) will come, which are interpreted based on statistical rules (Akoglu, 2018; Cohen, 1988a; Field, 2018; Nunnally & Bernstein, 1994; Salam, 2008). In sum up, the normality, outliers, linearity, independence, homoscedasticity and multicollinearity were checked by Skewness and Kurtosis, boxplot and Mahalanobis distance test, correlation, Durbin-Watson, scatterplot, and Tolerance and VIF test respectively of the data. The multiple linear regression's assumptions were all met. As a result, the researcher used a multiple linear regression for the study.

The calculated value of R, R-square significance F-value of the secondary mathematics teachers' in Kathmandu district are given in Table 7

Table 7

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.756[a]	.571	.550	.37458	.571	26.452	6	119	.000

a. Predictors: (Constant), WHS, RWC, LF, IM, JSR, PC

According to the analysis of the data in Table 7, there was a strongly positive correlation (relationship) (Cohen, 1988b) between the DV (job satisfaction) and the IVs (LF, PC, JSR, RWC, IM, and WHS) of mathematics teachers with $R = 0.76$, an $F (6, 119)=26.452$, a significance level of p=.000 < 0. 05 (α), and $R^2= 0.571$. This means that our model explains 57.1% of the variation in job satisfaction among secondary mathematics teachers by the IVs, which is statistically significant for model fit (Chatterjee & Hadi, 2012a; Nunnally & Bernstein, 1994).

Table 8

ANOVA[a]

	Sum of Squares	df	Mean Square	F	Sig.

1	Regression	22.268	6	3.711	26.452	.000[b]
	Residual	16.697	119	.140		
	Total	38.965	125			

a. Dependent Variable: JS

b. Predictors: (Constant), WHS, RWC, LF, IM, JSR, PC

Table 9

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig	Lower Bound	Upper Bound
1	(Constant)	-.173	.358		-.483	.630	-.882	.536
	LF	.203	.072	.193	2.832	.005	.061	.345
	PC	.262	.096	.291	2.726	.007	.072	.452
	JSR	.411	.107	.300	3.840	.000	.199	.622
	RWC	.149	.102	.120	1.461	.147	-.053	.350
	IM	.117	.088	.116	1.336	.184	-.057	.291
	WHS	-.055	.072	-.055	-.763	.447	-.196	.087

a. Dependent Variable: JS

According to the analysis of the data in Table 9, the regression coefficient between independent variable LF (Location and facilities) and dependent variable JS (Job satisfaction) was statistically significant, *r (119) = 0.203, t (119) = 2.832, p=0.005 ≤ 0.05,* at a confidence interval of *95% CI [0.061, 0.345],* so the job satisfaction of mathematics teachers increased by *0.203* for one unit increase of independent variable LF (Location and facilities). Similarly, the regression coefficient between independent variable PC (Policy and culture) and dependent variable JS (Job satisfaction) was statistically significant, *r (119) = 0.262, t (119) = 2.726, p=0.007 < 0.05,* at a confidence interval of *95% CI [0.072, 0.452],* so the job satisfaction of mathematics teachers increased by *0.262* for one unit increase of independent variable PC (Policy and culture).

The regression coefficient between independent variable JSR (Job security and recognition) and dependent variable JS (Job satisfaction) was statistically significant, *r (119) = 0.411, t (119) = 3.840, p=0.00 < 0.05,* at a confidence interval of *95% CI [0.199, 0.622],* so the job satisfaction of mathematics teachers increased by *0.411* for one unit increase of independent variable JSR (Job security and recognition). The regression coefficient between independent variable RWC (Relationship with co-workers) and dependent variable

JS (Job satisfaction) was not statistically significant, *r (119) = 0.*149, *t (119) = 1.461, p =0.147 > 0.05,* at a confidence interval of *95% CI [-0.053, 0.350],* so the job satisfaction of mathematics teachers increased by *0.149* for one unit increase of independent variable RWC (Relationship with co-workers) which was not found to be a significant change. The regression coefficient between independent variable IM (Instructional materials) and dependent variable JS (Job satisfaction) was not statistically significant, *r (119) = 0.*117, *t (119) = 1.336, p = 0.184 > 0.05,* at a confidence interval of *95% CI [-0.057, 0.291],* so the job satisfaction of mathematics teachers increased by *0.117* for one unit increase of independent variable IM (Instructional materials) which was not found to be a significant change.

The regression coefficient between independent variable WHS (Working hours and salary) and dependent variable JS (Job satisfaction) was not statistically significant, *r (119) = -0.055, t (119) = -0.763, p = 0.447 > 0.05,* at a confidence interval of *95% CI [-0.196, 0.87],* so the job satisfaction of mathematics teachers decreased by *-0.055* for one unit increase of independent variable WHS (Working hours and salary). Therefore, the independent variables: location and facilities, policy and culture, and job security and recognition show the positive impact among them job security and recognition is high impact. However, relationships with co-workers, instructional materials, and working hours and salary do not impact on the job satisfaction of secondary mathematics teachers.

Conclusion

The major goal of this study is to identify the prerequisites for using multiple linear regression. In this study, standardized questionnaires were administered to 126 secondary mathematics teachers in Kathmandu, and the results were evaluated using SPSS for linearity, normality, independence, outliers, homoscedasticity, and multicollinearity. The linearity, normality, independence, outliers, homoscedasticity and multicollinearity were checked by correlation, Skewness and Kurtosis, Durbin-Watson test, boxplot and Mahalanobis distance test, scatterplot, and Tolerance and VIF test respectively of the dataset. Multiple linear regression can be used to analyze a given dataset if all these pre-assumptions are satisfied. Multiple linear regression cannot be employed if one or more requirements are not satisfied. Therefore, before utilizing multiple linear regression, every researcher should ensure that these requirements are met, otherwise the result will be invalid or incorrect.

Finally, if the assumptions of a certain statistical approach are not satisfied, there are other non-parametric statistical procedures that researchers might use. Researchers should be conversant with them even if their power is frequently a little lower than that of parametric approaches since they provide useful options.

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). John Wiley & Sons, Kluwer Academic Publishers, Boston/Dordrecht/London. https://www.wiley.com/en-us/Outliers+in+Statistical+Data%2C+3rd+Edition-p-9780471930945

Belsley, D. A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression.* John Wiley & Sons, Inc., New York, NY,.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287–1294. https://ideas.repec.org/a/ecm/emetrp/v47y1979i5p1287-94.html

Chatterjee, S., & Hadi, A. S. (2012a). Regression analysis by example. In 5th (Ed.), *John Wiley and Sons,Inc*. www.it-ebooks.info

Chatterjee, S., & Hadi, A. S. (2012b). *Regression analysis by example.* shed by John Wiley & Sons, Inc., Hoboken, New. http://library.lol/main/F28583F14517B00AA7306D34F0348C53

Cohen, J. (1988a). Satatistical power analysis for the behavioral scieces. In *Lawrence Erlbaum Associates* (2nd ed.). chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Cohen, J. (1988b). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Cramer, D. (1998). *Fundamental statistics for social research*. TJ International Ltd, Padstow Cornwall. https://www.taylorfrancis.com/books/mono/10.4324/9780203360613/fundamental-statistics-social-research-duncan-cramer

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE Publications India Pvt Ltd. https://www.flipkart.com/discovering-statistics-using-spss/p/itmfeu8u8apsf8jq

Field, A. (2018). Discovering statistic using ibm spss statistic. In *Sage Edge* (5th ed.). Sage Publications Ltd. http://library.lol/main/75CCC20CC5CACEC5FB3B02FCE0DE6038

Gupta, S. C. (2011). *Fundamentals of statistics* (9th ed.). Himalaya Publishing House, New Delhi, India.

Houglin, D. C., & Lglewicz, B. (1987). Fine turning some resistant rules for outlier labeling. *Journal of American Statistical Association*, *82*, 1147–1149. https://doi.org/10.1080/0162 1459.1987.10478551

Meng, X. l., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172

Nunnally, J. . C., & Bernstein, I. H. (1994). Psychometric theory. In *Mcgraw-hill, Inc* (3rd ed.). http://library.lol/main/81BC7485BC968C17EA4672B6423B4ED9

Raykov, T., & Marcouldes, G. P. (2006). *A first course in structural equation modeling* (2nd ed.). New Jersey: Lawrence Erlbaum Asociate,Inc. https://www.routledge.com/A-First-Course-in-Structural-Equation-Modeling/Raykov-Marcoulides/p/book/9780805855883

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 13–14. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.nrc.gov/docs/ML1714/ML17143A100.pdf

Salam, A. (2008). *Interpreting Multiple Regression : A Short Overview*. 1–77.

Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, *45*(8), 1989–1996. https://doi.org/10.2307/1914122

Seawright, J. (2005). Qualitative comparative analysis vis-a-vis regression. *Studies in Comparative International Development*, *40*(1), 3–26. shorturl.at/dBRS6

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, *63*(324), 1343–1372. https://doi.org/10.2307/2285889

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*(3), 317–329. https://doi.org/10.1037/1082-989x.6.4.317

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Young, D. . S. (2017). *Handbook of regression methods*. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781315154701