សាកលវិទ្យាល័យ ភូមិន្ទភ្នំពេញ

ROYAL UNIVERSITY OF PHNOM PENH

DEPARTMENT OF INTERNATIONAL BUSINESS MANAGEMENT—IBM

APPLIED RESEARCH METHODOLOGY

BACHELOR PROGRAM

4TH YEAR, SEMESTER 1

MATERIAL INFO

Authors: Veasna SOU., Sambath PHOU., & Phichhang OU
Published online at www.spss.site: 13-Sept 2024
Revised version: 17-Dec-2024

TUTORING SERIES-07: Simple and Multiple Regression Analysis

SIMPLE REGRESSION

AND

MULTIPLE REGRESSION ANALYSIS

PURPOSES:

Simple and multiple regression analysis in SPSS is used to understand relationships between variables and make predictions based on those relationships (Groebner et al., 2018). Both analyses can be performed using the built-in regression functions, allowing users to easily input data, choose variables, and interpret results through output tables and graphs. Used in various fields such as social sciences, marketing, health research, and any discipline where understanding or predicting relationships among variables is important.

SIMPLE REGRESSION:

In simple linear regression, each observation consists of two variables: one for the independent variable and one for the dependent variable (Anderson et al., 2014).

- To assess the relationship between one independent variable (predictor) and one dependent variable (outcome).
- **Key Uses:**
  - Predicting the value of the dependent variable based on the independent variable.
  - Understanding the strength and direction of the relationship (positive or negative).
  - Calculating the effect size and assessing the significance of the predictor variable.
- **Rules and Guidelines:**
  - **Sample Size:** A minimum of 15-20 observations for each predictor variable is often recommended. However, more data is preferable to achieve reliable results.
  - **Linearity:** The relationship between the independent and dependent variable should be linear. This can be visually assessed using scatter plots.
  - **Normality:** Residuals (errors) should be approximately normally distributed, especially for inference tests (like t-tests on regression coefficients).
  - **Homoscedasticity:** The variance of residuals should be constant across all levels of the independent variable(s). This means that the spread of residuals should not increase or decrease as predicted values increase.
  - **No Multicollinearity:** Since there is only one predictor in simple regression, this rule does not directly apply, but you should ensure the predictor is not perfectly correlated with another variable if it's included in a subsequent analysis.

MULTIPLE REGRESSION:

Multiple regression analysis is the study of how a dependent variable y is related to two or more independent variables (Anderson et al., 2014).

- To examine the relationship between one dependent variable and multiple independent variables.
- Key Uses:
  - Identifying the most impactful predictors on the dependent variable.
  - Evaluating how each independent variable contributes to the prediction of the dependent variable, adjusting for the effects of other predictors.
  - Understanding interaction effects among independent variables.
  - Conducting hypothesis tests and obtaining confidence intervals for the regression coefficients.
- **Rules and Guidelines:**

- **Sample Size:** Ideally, at least 10-15 observations per predictor variable in the model. However, more data can help increase the robustness of the model.
- **Multicollinearity:** Check for high correlations among independent variables. Variance Inflation Factor (VIF) values greater than 10 may indicate problematic multicollinearity.
- **Linearity:** The relationship between each independent variable and the dependent variable should remain linear. Consider transformations if necessary.
- **Normality of Residuals:** The residuals should be normally distributed. This is particularly important for hypothesis testing on regression coefficients.
- **Homoscedasticity:** Residuals should exhibit constant variance. Assess this with a plot of residuals versus predicted values.
- **Independence of Residuals:** The observations should be independent of one another, especially in time series data or grouped data scenarios.

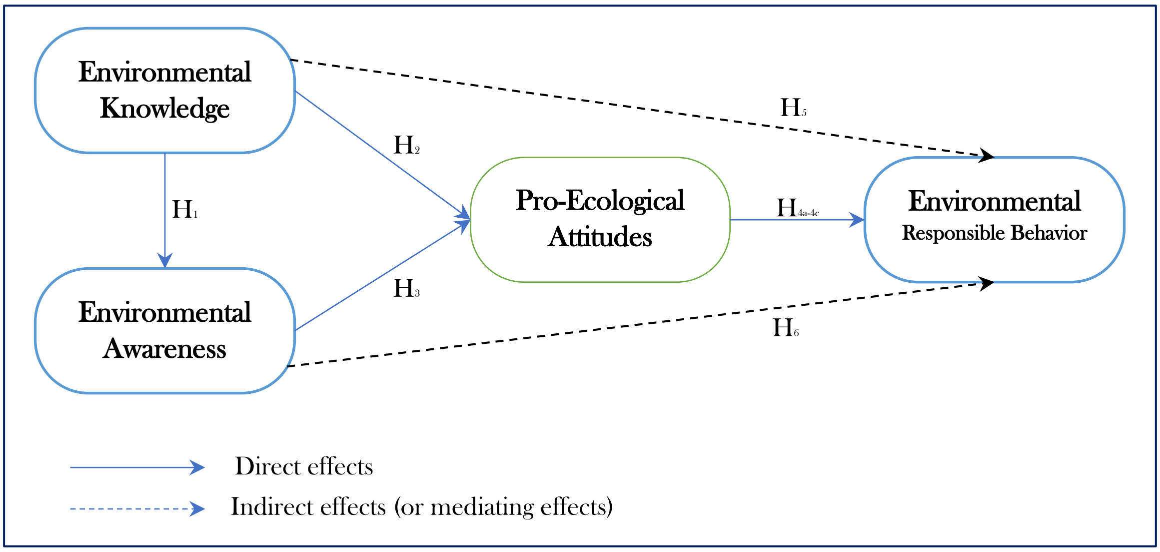
These rules serve as guidelines to help ensure that the regression analyses yield valid and interpretable results.

1. CONCEPTUAL FRAMEWORK AND HYPOTHESES DEVELOPMENT:

This conceptual framework has key roles as follow:

Table 1. Summary of Hypotheses Development

Research Constructs	Independent Variables (X)	Mediating Variables (MEV)	Dependent Variables (Y)
Direct Effects			
H_1: ENK→ENA	Environmental Knowledge		Environmental Awareness
H_2: ENK→PEA	Environmental Knowledge		Pro-Ecological Attitudes
H_3: ENA→PEA	Environmental Awareness		Pro-Ecological Attitudes
H_4: PEA →ENR	Pro-Ecological Attitudes		Environmentally Responsible Behavior
Indirect Effects			
H_5: ENK→ PEA → ENR	Environmental Knowledge	Pro-Ecological Attitudes	Environmentally Responsible Behavior
H_4: ENA → PEA → ENR	Environmental Awareness	Pro-Ecological Attitudes	Environmentally Responsible Behavior



2. RULE OF THUMBS: REGRESSION ANALYSIS

Table 2. The Rule of Thumbs: Regression Analysis

Criterion	Threshold Values
1. R^2 (R-square)	≥ 0.10 (10%)
2. Adjusted- R^2	≥ 0.10 (10%)
3. F-value	≥ 4
4. t-value	≥ |1.96|
5. p-value	< 0.05
6. Durbin-Watson (D-W)	[1.50-2.50]
7. VIF	≤ 2.5

Sources: (i.e., Hair Jr et al., 2019; Hair Jr et al., 2021; Johnston et al., 2018)

3. EQUATIONS OF SIMPLE AND MULTIPLE REGRESSION ANALYSIS

3.1. EQUATION OF SIMPLE REGRESSION

An equation of simple regression as written below:

$Y_1 = \boldsymbol{a_1} X_1 + \boldsymbol{b} + \boldsymbol{\varepsilon}_0$ or $Y_1 = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} X_1 + \boldsymbol{\varepsilon}_0$ (Equation 1.1)

Where:

- $\boldsymbol{\beta_1}$: *is a slope of regression line (or curve) direction*
- $\boldsymbol{\beta_0}$: *is an intercept (or constant value)*
- ***X_1**: is an independent variable*
- ***Y_1**: is a dependent variable*
- $\boldsymbol{\varepsilon}$: *is the standardized error or random error term*

Then, $Y_1 = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} X_1 + \boldsymbol{\varepsilon}_0$ (Equation 1.2)

Example 1:

In the above conceptual framework, our research hypothesis (H_1) is to predict the relationship between "Environmental Knowledge" and "Environmental Awareness". Therefore, the following equation of simple regression will be:

$\boldsymbol{ENA} = \beta_0 + \boldsymbol{\beta_1 ENK} + \boldsymbol{\varepsilon}_0$ (Equation 1.3)

Where:

- $\boldsymbol{\beta_1}$: *is a slope of regression line (or curve) direction*
- $\boldsymbol{\beta_0}$: *is an intercept (or constant value)*
- ***X_1**: is Environmental Knowledge: ENK (Independent Variable)*
- ***Y_1**: is Environmental Awareness: ENA (Dependent Variable)*
- $\boldsymbol{\varepsilon}$: *is the standardized error or random error term*

3.2. EQUATION OF MULTIPLE REGRESSION

An equation of multiple regression as written below:

$Y_1 = \boldsymbol{b} + (\boldsymbol{a}_1 X_1 + \boldsymbol{a}_2 X_2 + \cdots + \boldsymbol{a}_n X_n) + \boldsymbol{\varepsilon}_0$ or $Y_1 = \boldsymbol{\beta_0} + (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n) + \boldsymbol{\varepsilon}_0$ (Equation 2.1)

Where:

- β_1: *is a slope of regression line (or curve) direction*
- β_0: *is an intercept (or constant value)*

- $X_{1\text{-}n}$ *: is the independent variables*
- Y_1 *: is a dependent variable*
- $\boldsymbol{\varepsilon}$*: is the standardized error or random error term*

Then, $Y_1 = \boldsymbol{\beta_0} + (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n) + \boldsymbol{\varepsilon_0}$ (Equation 2.2)

Example 2:

In the above conceptual framework, our research hypotheses (H_2) and (H_3) are to predict the relationship between "Environmental Knowledge" and "Environmental Awareness" on "Pro-Environmental Attitude". Therefore, the following equation of simple regression will be:

$\mathbf{PEA} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 ENK} + \boldsymbol{\beta_2 ENA} + \boldsymbol{\varepsilon_0}$ (Equation 2.3)

Where:

- $\boldsymbol{\beta_1}$*: is a slope of regression line (or curve) direction of relationship for Hypothesis 1*
- $\boldsymbol{\beta_2}$*: is a slope of regression line (or curve) direction of relationship for Hypothesis 2*
- $\boldsymbol{\beta_0}$*: is an intercept (or constant value)*
- X_1*: is Environmental Knowledge: ENK (Independent Variable)*
- X_2*: is Environmental Awareness: ENA (Independent Variable)*
- Y_1*: is Pro-Environmental Attitude: PEA (Dependent Variable)*
- $\boldsymbol{\varepsilon}$*: is the standardized error or random error term*

Overall: The SPSS process must illustrate the slope and constant value to test the relationship between the independent variable (X) and the independent variable (Y).

4. STEP BY STEP...

PART I. SIMPLE REGRESSION ANALYSIS FOR A SINGLE FACTOR

Go to **Analyze** >> **Regression** >> **Linear** (Figure 1.1) **>>** [move the mean score of "Environmental Knowledge—ENKX" to **"Block box = Independent Variable(s) box"** and mean score of "Environmental Awareness—ENAX" must move to "Dependent Variable" box (Figure 1.2) and **>> Statistics (**check: R-square change, Part and partial correlations, Collinearity diagnostics, Confident Interval (95%), and Durbin-Watson**)** and rest o of other function just let it defaults, then click **Continue** (Figure 1.2) **>> OK.** Then, you will see the following outputs of simple regression analysis for the research hypothesis (H1). Refer to the outputs (Figure 1.3) and shown in (Table 3) of the H1 below:

Hypothesis (H1): Environmental Knowledge has a positive influence on Environmental Awareness.

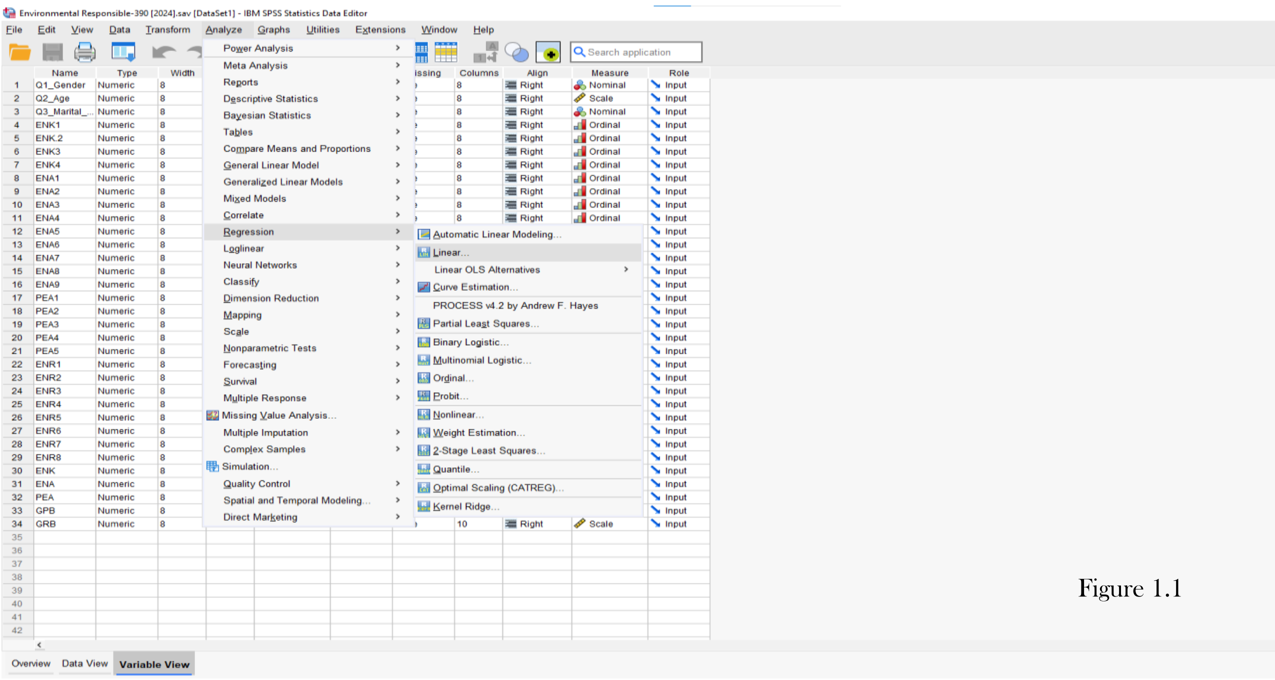


Figure 1.1

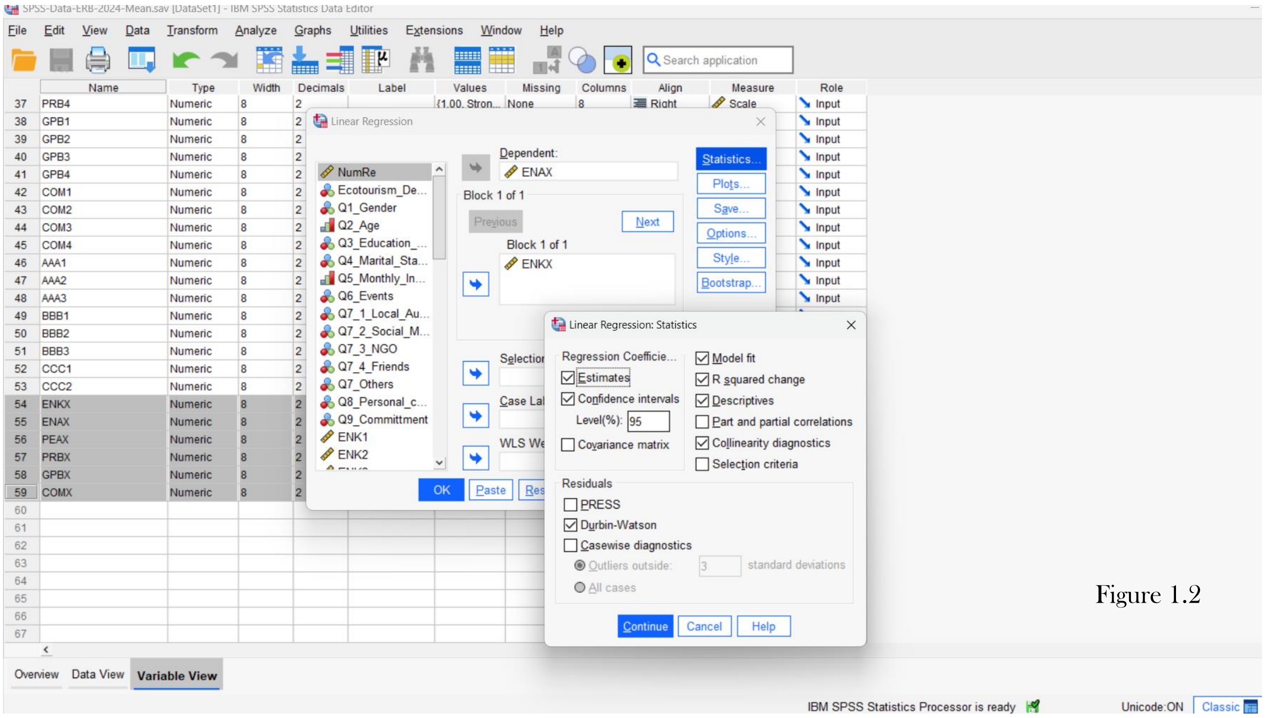
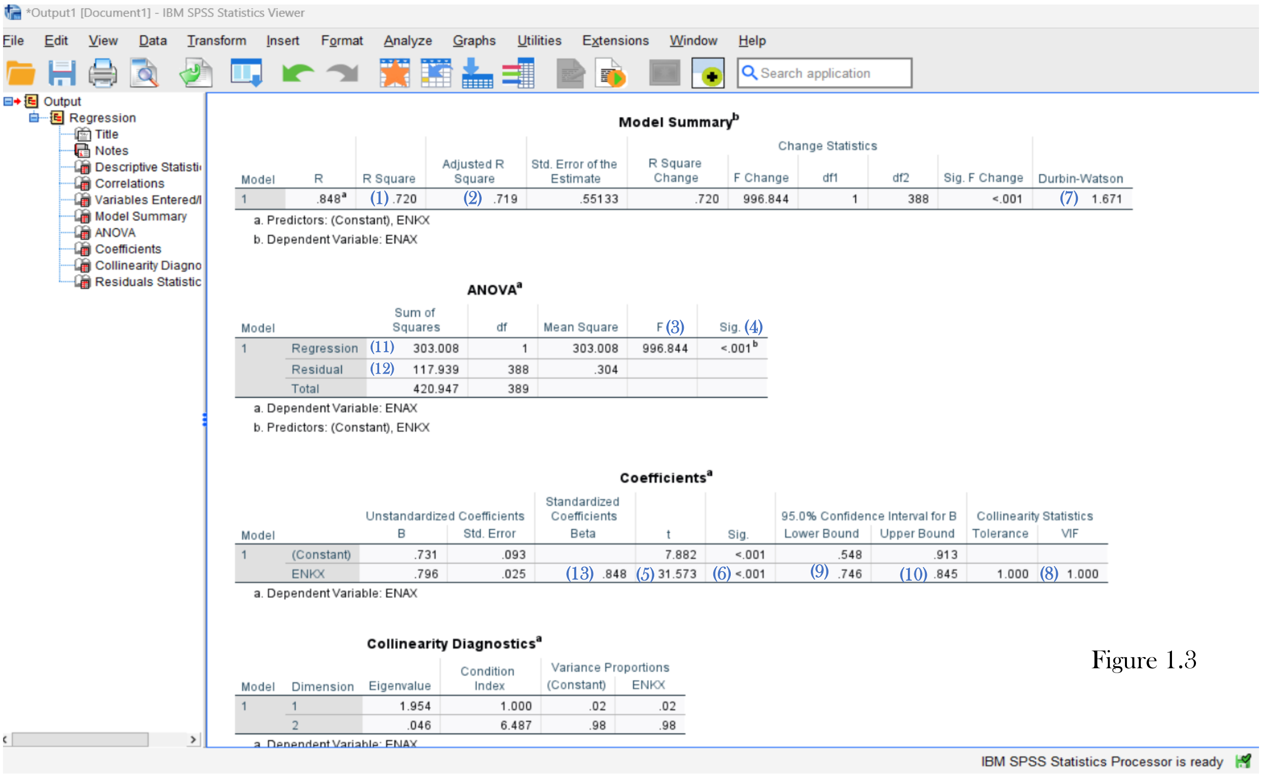


Figure 1.2



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.848[a]	(1) .720	(2) .719	.55133	.720	996.844	1	388	<.001	(7) 1.671

a. Predictors: (Constant), ENKX
b. Dependent Variable: ENAX

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F (3)	Sig. (4)
1	Regression	(11) 303.008	1	303.008	996.844	<.001[b]
	Residual	(12) 117.939	388	.304		
	Total	420.947	389			

a. Dependent Variable: ENAX
b. Predictors: (Constant), ENKX

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	Upper Bound	Collinearity Statistics Tolerance	VIF
1	(Constant)	.731	.093		7.882	<.001	.548	.913		
	ENKX	.796	.025	(13) .848	(5) 31.573	(6) <.001	(9) .746	(10) .845	1.000	(8) 1.000

a. Dependent Variable: ENAX

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	ENKX
1	1	1.954	1.000	.02	.02
	2	.046	6.487	.98	.98

a. Dependent Variable: ENAX

Figure 1.3

Table 3. The Result of Hypothesis (H1)

Independent Variable	Dependent Variable: Environmentally Responsible Behavior(Y_1)
	β_1 (Model-1)
Environmental Knowledge (X_1)	(13) 0.848***
R^2 (R-square)	(1) 0.720
Adjusted-R^2	(2) 0.719
F-value (Significant of p-value)	(3) 996.844 (p<0.001) (4)
t-value	(5) 31.573
p-value	(6) <0.001
Durbin-Watson (D-W)	(7) 1.671
VIF	(8) 1.00
Confident Interval (CI)	(9) [0.746 - 0.845] (10)
D.F (Regression)	(11) 1
D.F (Residual)	(12) 388

*Note: ***p<0.001, **p<0.05, *p<0.10 and significant at t-value > |1.96|. d.f = degree of freedom*

In Table 3, simple linear regression analysis was conducted to evaluate the extent to which "Environmental Knowledge" could predict "Environmental Awareness" of local tourism community. A significant regression was found F (1,388) = 996.844, p = < 0.001). The R^2 was 0.720, indicating "Environmental Knowledge—ENK" explained approximately 72.0% of the variance in "Environmental Awareness—ENA". The simple regression equation in Model-1 was:

$$Y_1(ENK) = \beta_0 + \beta_1 X_1(ENA) + \varepsilon_0 \qquad \textit{(Equation 1)}$$

$$\textit{Then: } Y_1(ENK) = 0.731 + 0.848PEA + 0.093 \qquad \textit{(Equation 1)}$$

That is, for increase in "Environmental Knowledge", the predicted "Environmental Awareness" increased by approximately 0.746 or 74.6%. Confidence intervals indicated that we can be 95% certain that the slope to predict "Environmental Awareness" from "Environmental Knowledge "is between 0.746 and 0.845. The regression coefficients showed that for "Environmental Awareness" increased by an average of 0.848(84.8%) (β_1 = 0.848, SE = 0.093, t = 31.573, p < 0.001). This result indicated that "Environmental Knowledge" are a significant predictor of "Environmental Awareness", supporting the hypothesis that increased "Environmental Knowledge" is associated with "Environmental Awareness". Overall, the findings suggest that encouraging local tourism community to engage in more "Environmental Knowledge" could enhance their "Environmental Awareness".

PART II. SIMPLE REGRESSION ANALYSIS FOR A MULTIPLE FACTORS

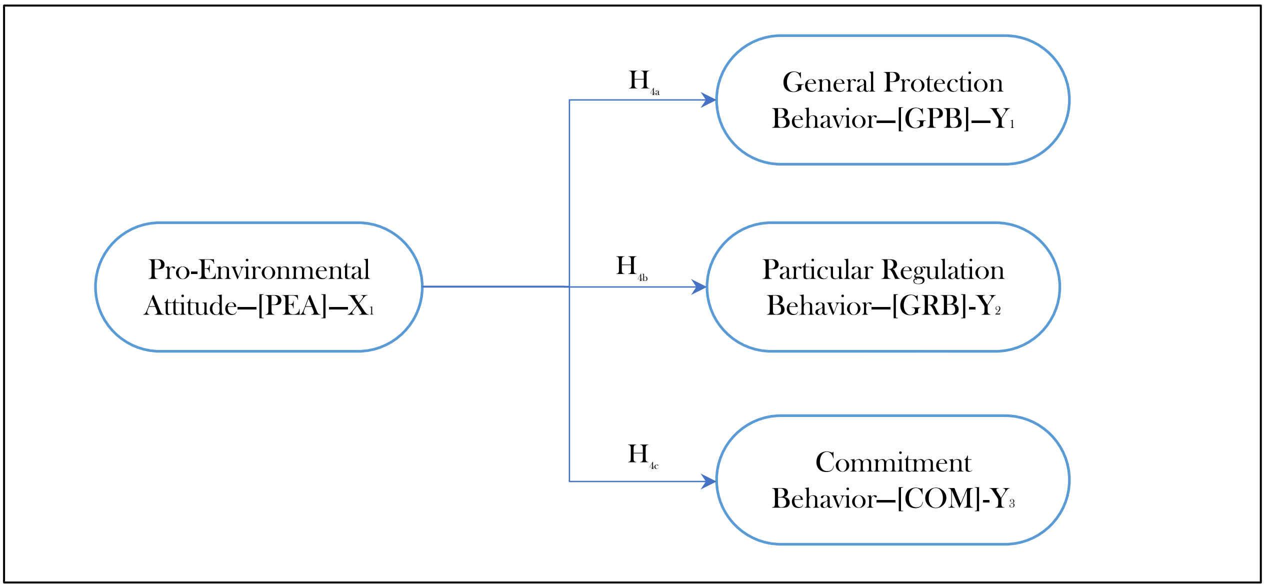
Go to **Analyze** >> **Regression** >> **Linear** (Figure 2.1) **>>** [move the mean score of "Pro-Environmental Attitude—PEA" to **"Block box = Independent Variable(s) box"** and mean score of Environmental Responsibility—(i.e., GPB, GRB, and COM) must move to "Dependent Variable" box (Figure 2.2) and **>> Statistics (**check: R-square change, Part and partial correlations, Collinearity diagnostics, Confident Interval (95%), and Durbin-Watson**)** and rest o of other function just let it defaults, then click **Continue** (Figure 2.2) **>> OK.** Then, you will see the following outputs of simple regression analysis for the research hypothesis ($H_{4a\text{-}4c}$). Refer to the outputs (shown in Table 4) of the H4a, H4b, and H4c below:

ATTENTION: In this case, a variable of "Environmentally Responsible Behavior "consists of three sub-dimensions (1)- General Protection Behavior—GPB, (2)-Particular Regulation Behavior—GRB, and (3)-Commitment Behavior—COM. We already computed mean score of these factors in SPSS data set. Therefore, we will have three simple regression results with Hypothesis (H_{4a}), (H_{4b}), and (H_{4c}). Let's break down the relationship of this hypothesis below. Based on this break down relationships, we need to run the simple regression twice. First, to test the relationship of H_{4a}. Second, to the test the relationship of H_{4b}, and third, test the relationship of H_{4c}

Hypothesis (H4a): "Pro-Environmental Attitude" has a positive influence on "General Protection Behavior."

Hypothesis (H4b): "Pro-Environmental Attitude" has a positive influence on "Particular Regulation Behavior."

Hypothesis (H4c): "Pro-Environmental Attitude" has a positive influence on "Commitment Behavior."



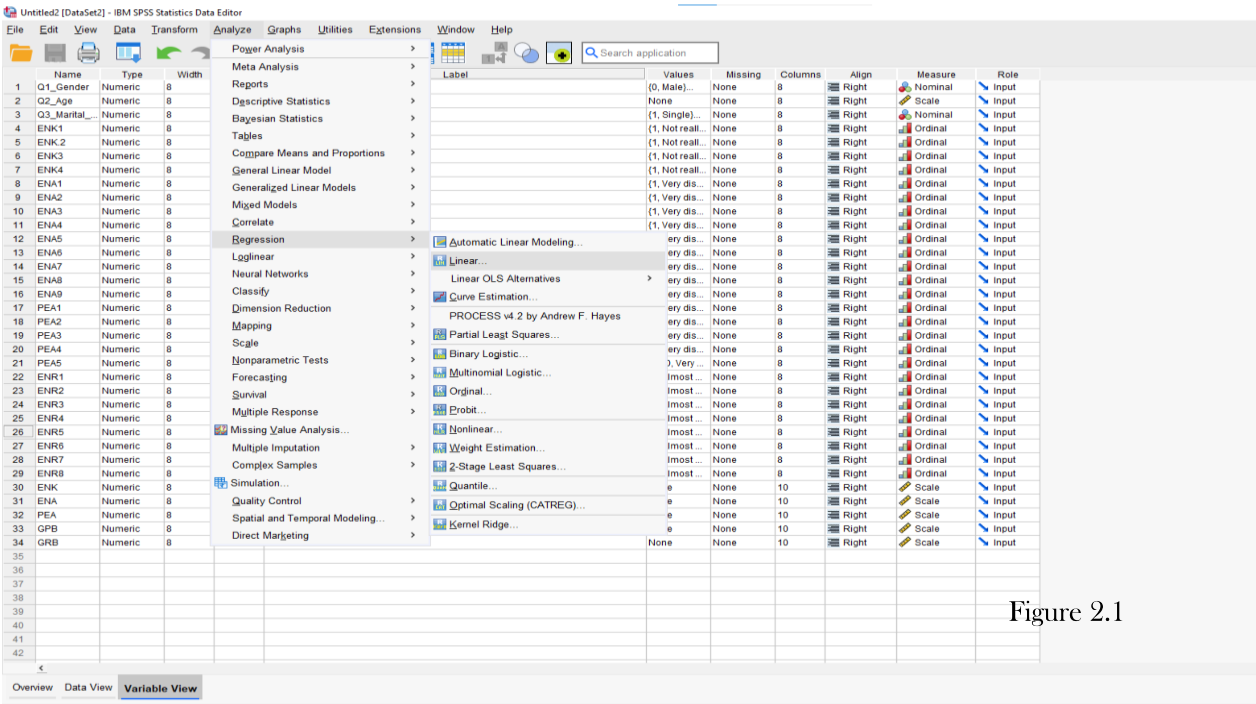


Figure 2.1

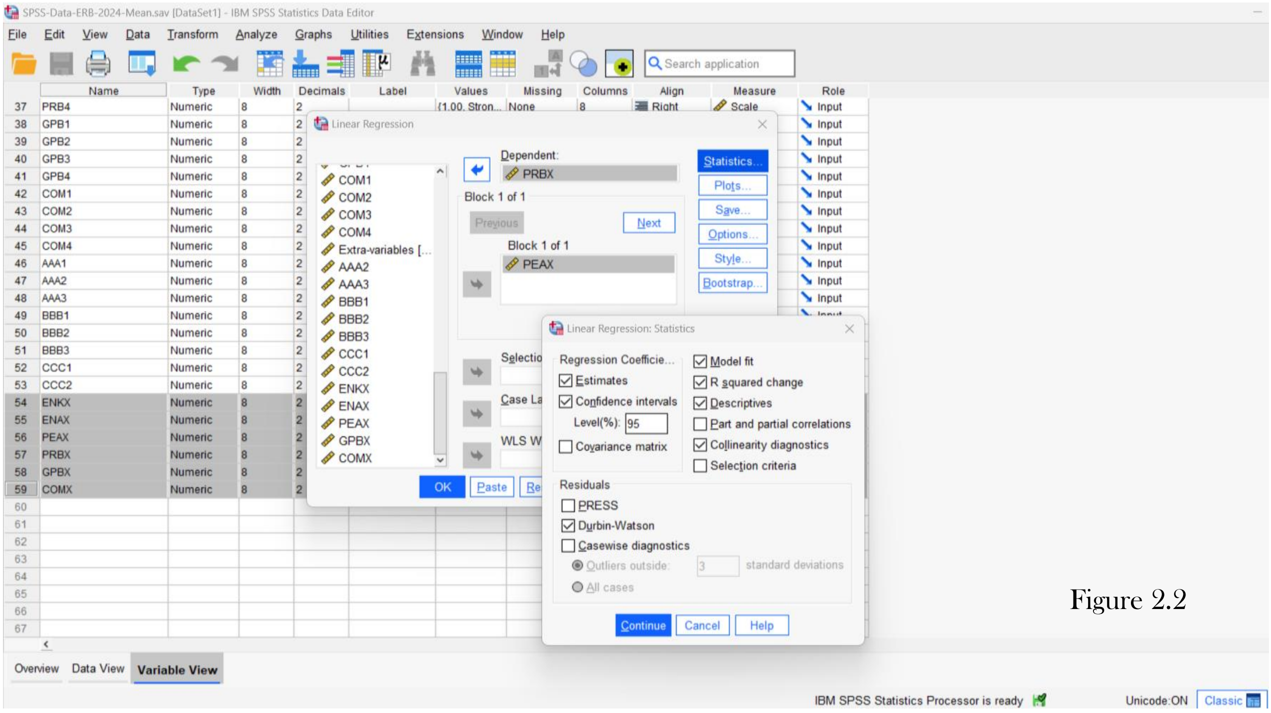


Figure 2.2

4.1. Simple Regression: Outputs

4.1.1. Simple Regression for Hypothesis (H4a)

Model Summary[b]

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.570[a]	(1) .324	(2) .323	.75159	.324	186.280	1	388	<.001	(7) 1.624

a. Predictors: (Constant), PEAX

b. Dependent Variable: PRBX

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	(11) 105.227	1	105.227	(3) 186.280	(4) <.001[b]
	Residual	(12) 219.175	388	.565		
	Total	324.401	389			

a. Dependent Variable: PRBX

b. Predictors: (Constant), PEAX

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	1.755	.133		13.222	<.001	1.494	2.016		
	PEAX	.532	.039	(13) .570	(5) 13.648	(6) <.001	(9) .456	(10) .609	1.000	(8) 1.000

a. Dependent Variable: PRBX

4.1.2. *Simple Regression: Outputs—For Hypothesis (H4b)*

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	.691[a]	.478	.477	.70628	.478	355.104	1	388	<.001	1.507

a. Predictors: (Constant), PEAX

b. Dependent Variable: GPBX

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	177.138	1	177.138	355.104	<.001[b]
	Residual	193.548	388	.499		
	Total	370.686	389			

a. Dependent Variable: GPBX

b. Predictors: (Constant), PEAX

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	1.261	.125		10.111	<.001	1.016	1.506		
	PEAX	.690	.037	.691	18.844	<.001	.618	.763	1.000	1.000

a. Dependent Variable: GPBX

4.1.3. Simple Regression: Outputs—For Hypothesis (H4c)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.638[a]	.407	.406	.76664	.407	266.787	1	388	<.001	1.578

a. Predictors: (Constant), PEAX
b. Dependent Variable: COMX

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	156.800	1	156.800	266.787	<.001[b]
	Residual	228.040	388	.588		
	Total	384.840	389			

a. Dependent Variable: COMX
b. Predictors: (Constant), PEAX

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	Upper Bound	Collinearity Statistics Tolerance	VIF
1	(Constant)	1.337	.135		9.873	<.001	1.070	1.603		
	PEAX	.650	.040	.638	16.334	<.001	.571	.728	1.000	1.000

a. Dependent Variable: COMX

4.2. *TABLE FORMAT OF SIMPLE REGRESSION*

Table 4. The Result of Hypothesis H4a, H4b, and H4c

Independent variable	Dependent variable: Environmentally Responsible Behavior		
	GPB (Y_1)	GRB (Y_2)	COM (Y_3)
	β_1 (Model-1)	β_2 (Model-2)	β_3 (Model-3)
Pro-Environmental Attitudes (X_1)	(13) 0.570***	0.691***	0.638***
R^2 (R-square)	(1) 0.324	0.478	0.407
Adjusted- R^2	(2) 0.323	0.477	0.406
F-value (Significant of p-value)	(3) 186.280 (p<0.001) (4)	355.104 (p<0.001)	266.787 (p<0.001)
t-value	(5) 13.648	18.844	16.334
p-value	(6) <0.001	<0.001	<0.001
Durbin-Watson (D-W)	(7) 1.624	1.507	1.578
VIF	(8) 1.00	1.00	1.00
Confident Interval (CI)	(9) [0.456 - 0.609] (10)	[0.618 - 0.763]	[0.571-0.728]
D.F (Regression)	(11) 1	1	1
D.F (Residual)	(12) 388	388	389

*Note: ***p<0.001, **p<0.05, *p<0.10 and significant at t-value > |1.96|*

4.3. RESULTS AND INTERPRETATION

4.3.1 For Hypothesis (H4a)

In Table 4, simple linear regression analysis was conducted to evaluate the extent to which "Pro-Environmental Attitudes" could predict "Green Protected Behavior" of local tourism community. A significant regression was found $F(1,388) = 186.280$, $p < 0.001$). The R^2 was 0.324, indicating "Pro-Environmental Attitudes" explained approximately 32.4% of the variance in "Green Protected Behavior". The simple regression equation in Model-1 was:

$$Y_1(GPB) = \beta_0 + \beta_1 X_1(PEA) + \varepsilon_0 \qquad \textit{(Equation 2)}$$
$$\textit{Then: } Y_1(GPB) = 1.755 + 0.570PEA + 0.133 \qquad \textit{(Equation 2)}$$

That is, for increase in "*Pro-Environmental Attitudes*", the predicted "*Green Protected Behavior*" increased by approximately 0.456 or 45.6%. Confidence intervals indicated that we can be 95% certain that the slope to predict "*Green Protected Behavior*" from "*Pro-Environmental Attitudes* "is between 0.456 and 0.609. The regression coefficients showed that for "Green Protected Behavior" increased by an average of 0.570 (57.0%) (β_1 = 0.570, SE = 0.133, t = 13.648, p < 0.001). This result indicated that "Pro-Environmental Attitudes" are a significant predictor of "Green Responsible Behavior", supporting the hypothesis that increased "Pro-Environmental Attitudes" is associated with "Green Responsible Behavior". Overall, the findings suggest that encouraging local tourism community to engage in more "Pro-Environmental Attitudes" could enhance their "Green Protected Behavior".

4.3.2 For Hypothesis (H4b)

In Table 4, simple linear regression analysis was conducted to evaluate the extent to which "Pro-Environmental Attitudes" could predict "Green Responsible Behavior" of local tourism community. A significant regression was found F (1,388) = 355.104, p = < 0.001). The R^2 was 0.478, indicating "Pro-Environmental Attitudes" explained approximately 47.8% of the variance in "Green Responsible Behavior". The simple regression equation in Model-2 was:

$$Y_2(GRB) = \beta_0 + \beta_2 X_1(PEA) + \varepsilon_0 \qquad \textit{(Equation 3)}$$
$$\textit{Then: } Y_2(GRB) = 1.261 + 0.691PEA + 0.125 \qquad \textit{(Equation 3)}$$

That is, for increase in "*Pro-Environmental Attitudes*", the predicted "Green Responsible Behavior" increased by approximately 0.618 or 61.8%. Confidence intervals indicated that we can be 95% certain that the slope to predict "*Green Protected Behavior*" from "*Pro-Environmental Attitudes* "is between 0.618 and 0.763. The regression coefficients showed that for "Green Responsible Behavior" increased by an average of 0.691 (69.1%) (β_1 = 0.691, SE = 0.125, t = 18.844, p < 0.001). This result indicated that "Pro-Environmental Attitudes" are a significant predictor of "Green Responsible Behavior", supporting the hypothesis that increased "Pro-Environmental Attitudes" is associated with "Green Responsible Behavior". Overall, the findings suggest that encouraging local tourism community to engage in more "Pro-Environmental Attitudes" could enhance their "Green Responsible Behavior". Therefore, the results show a very weak relationship between "pro-environmental attitudes" and "green responsible behavior," with an 8.1% correlation. This means that among 390 participants, there are only 186 respondents (i.e., [390x0.478] = 186) who have the concepts of "green responsible behavior" in their local tourism community.

4.3.3 For Hypothesis (H4c)

In Table 4, simple linear regression analysis was conducted to evaluate the extent to which "Pro-Environmental Attitudes" could predict "Commitment Behavior" of local tourism community. A significant regression was found F (1,389) = 266.787, p = < 0.001). The R^2 was 0.407, indicating "Pro-Environmental Attitudes" explained approximately 40.7% of the variance in "Commitment Behavior". The simple regression equation in Model-3 was:

$$Y_3(COM) = \beta_0 + \beta_3 X_1(PEA) + \varepsilon_0 \qquad \textit{(Equation 4)}$$
$$\textit{Then: } Y_3(COM) = 1.337 + 0.638PEA + 0.135 \qquad \textit{(Equation 4)}$$

That is, for increase in "*Pro-Environmental Attitudes*", the predicted "Green Responsible Behavior" increased by approximately 0.571 or 57.1%. Confidence intervals indicated that we can be 95% certain that the slope to predict "*Green Protected Behavior*" from "*Pro-Environmental Attitudes* "is between 0.571 and 0.728. The regression coefficients showed that for "Green Responsible Behavior" increased by an average of 0.638 (63.8%)

(β_1 = 0.638, SE = 0.135, t = 16.334, p < 0.001). This result indicated that "Pro-Environmental Attitudes" are a significant predictor of "Commitment Behavior", supporting the hypothesis that increased "Pro-Environmental Attitudes" is associated with "Commitment Behavior". Overall, the findings suggest that encouraging local tourism community to engage in more "Pro-Environmental Attitudes" could enhance their "Commitment Behavior". Therefore, the results show a moderate relationship between "pro-environmental attitudes" and "Commitment Behavior," with an 40.7% correlation. This means that among 390 participants, there are only 159 respondents (i.e., [390x0.407] = 159) who have the concepts of "Commitment Behavior" in their local tourism community.

In summary, all sub-dimensions of "Environmentally Responsible Behavior" are well-supported by the prediction of a key independent research variable of "Pro-Ecological Attitudes". Therefore, H4a, H4b, and H4c are accepted.

PART III. MULTIPLE REGRESSION

In conceptual model, we have one multiple linear regression for Hypothesis 2 and Hypothesis 3. The following step-by-step is provided below:

Go to **Analyze** >> **Regression** >> **Linear** (Figure 3.1) **>>** [move the mean scores of Environmental Knowledge—(ENK) and Environmental Awareness—(ENA) to **"Block box = Independent Variable(s) box"** (i.e., ENK for Block 1 and ENA for Block 2 by clicking on next) and mean score of "Pro-Environmental Attitude—PEA" must move to "Dependent Variable" box (Figure 3.2) and **>> Statistics** (check: R-square change, Part and partial correlations, Collinearity diagnostics, Confident Interval (95%), and Durbin-Watson) and rest o of other function just let it defaults, then click **Continue** (Figure 3.2) **>> OK.** Then, you will see the following outputs of multiple linear regression for the research hypotheses (H_2) and (H_3). Refer to the outputs (shown in Table 5) of the H_2 and H_3 below:

This study proposes the research hypotheses, as followed:

Hypothesis (H_2): "Environmental Knowledge" has a positive influence on "Pro-Environmental Attitude."

Hypothesis (H_3): "Environmental Awareness" has a positive influence on "Pro-Environmental Attitude."

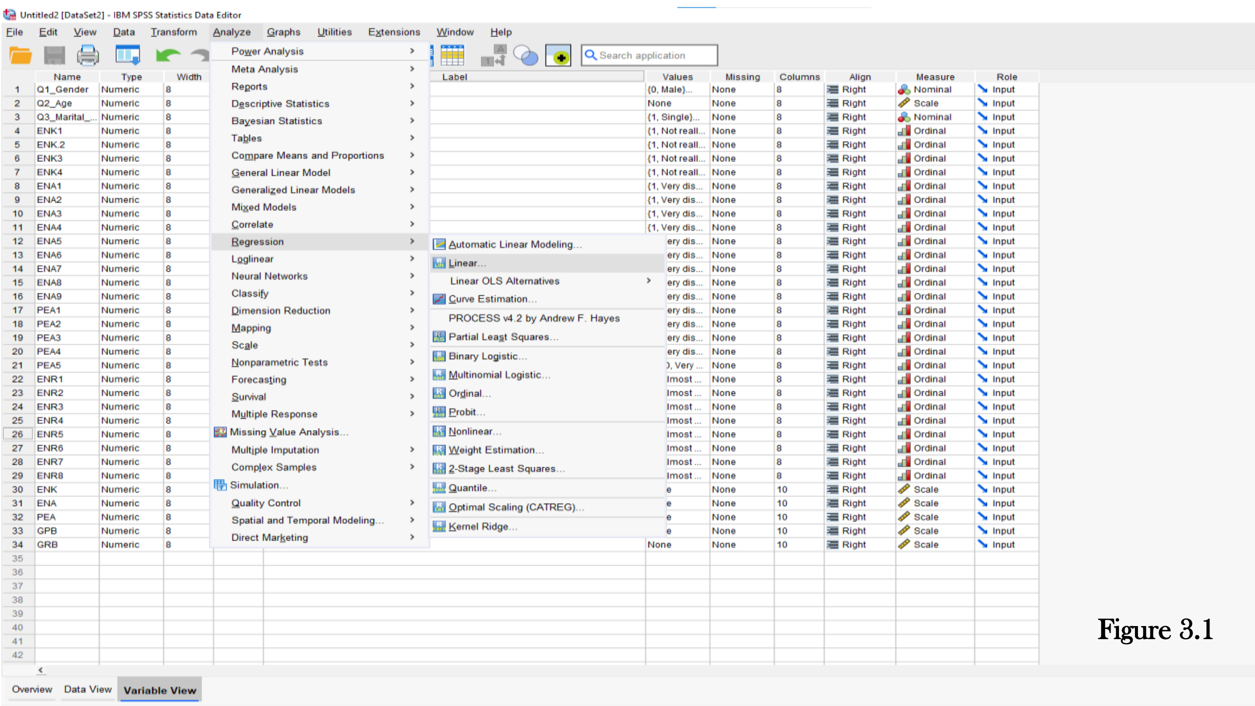


Figure 3.1

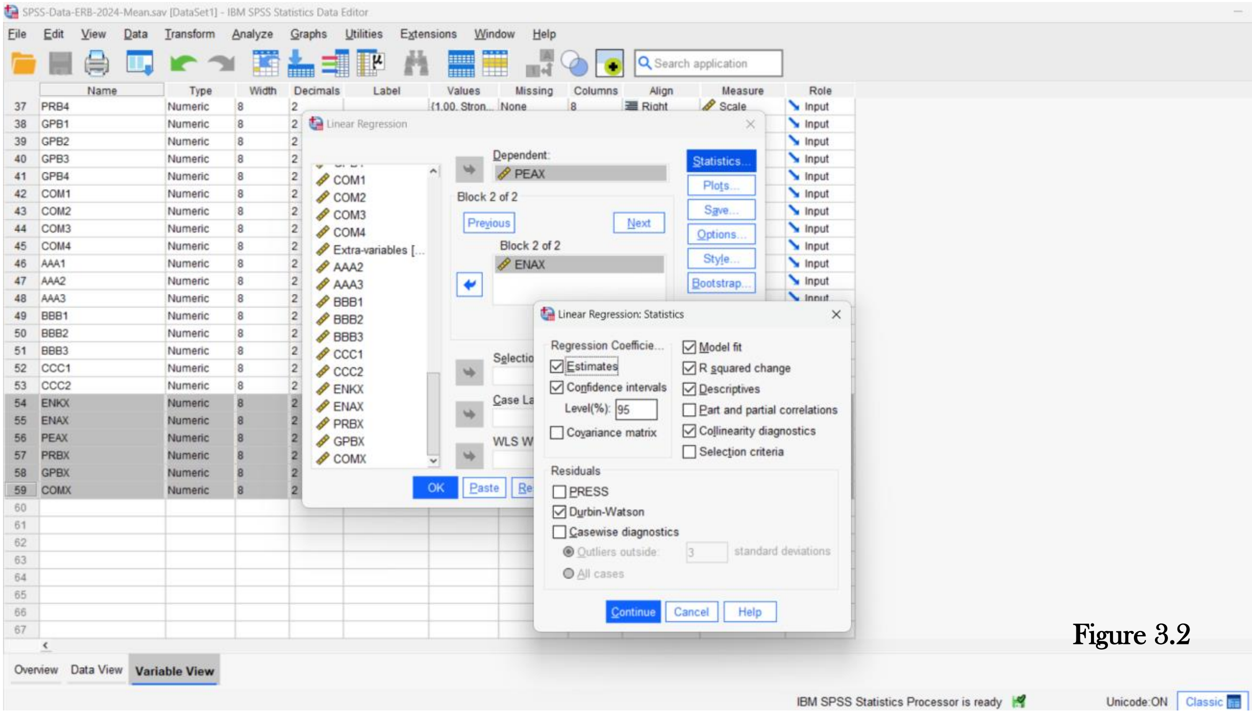


Figure 3.2

Table 5. The Result of Hypothesis (H_2) and (H_3)

Independent variables	**Dependent variable: Pro-Environmental Attitudes–(Y_1)**	
	β_1 (Model-1)	β_2 (Model-2)
Environmental Knowledge (X_1)	(13) 0.295**	
Environmental Awareness (X_2)		(M) 0.441***
R^2 (R-square)	(1) 0.448	(A) 0.502
Adjusted- R^2	(2) 0.446	(B) 0.499
F-value (Significant of p-value)	(3) 314.284 (p<0.001) (4)	(C) 194.971 (p<0.001) (D)
t-value	(5) 4.355	(E) 6.500
p-value	(6) <0.001	(F) <0.001
Durbin-Watson (D-W)	(7) 1.510	(G) 1.510
VIF	(8) 3.569	(H) 3.569
Confident Interval (CI)	(9) [0.143 – 0.377] (10)	(I) [0.289 – 0.539] (J)
d.f (Regression)	(11) 1	(K) 2
d.f (Residual)	(12) 388	(L) 387

*Note: ***p<0.001, **p<0.05, *p<0.10 and significant at t-value > |1.96|, d.f= degree of freedom*

The interpretation of multiple regression is similar to simple regression because simple regression is a special case of multiple regression (Doane & Seward, 2016). In Table 5, multiple linear regression analysis was conducted to evaluate the extent to which "Environmental Knowledge" and "Environmental Awareness" could predict "Pro-Environmental Attitudes" of local tourism community. In Model-1, a significant regression was found F (1,388) = 314.284, p = < 0.001). The R^2 was 0.448, indicating "Environment Knowledge" explained approximately 44.8% of the variance in "Pro-Environmental Attitudes". In Model-2, a significant regression was found F (2,387) = 194.971, p = < 0.001). The R^2 was 0.502, indicating "Environment Awareness" explained approximately 50.2% of the variance in "Pro-Environmental Attitudes". Thus, the multiple regression equation in was:

$$Y_1(EPA) = \beta_0 + \beta_1 X_1(ENK) + \beta_2 X_2(ENA) + \varepsilon_0 \qquad \textit{(Equation 5)}$$

Then: $$Y_1(EPA) = 0.891 + 0.295ENK + 0.441ENA + 0.125 \qquad \textit{(Equation 5)}$$

That is, for increase in "*Environmental Knowledge"*, the predicted "*Pro-Environmental Attitudes*" increased by approximately 0.143 or 14.3%. Confidence intervals indicated that we can be 95% certain that the slope to predict "*Environmental Knowledge*" from "*Pro-Environmental Attitudes* "is between 0.143 and 0.377. The regression coefficients showed that for "*Pro-Environmental Attitudes*" increased by an average of 0.446 (44.6%) (β_1 = 0.295, SE = 0.060, t = 4.355, p <0.001). This result indicated that "Pro-Environmental Attitudes" are a significant predictor of "*Environmental Knowledge*", supporting the hypothesis that increased "*Environmental Knowledge*" is associated with "Pro-Environmental Attitudes". Overall, the findings suggest that encouraging local tourism community to engage in more "*Environmental Knowledge*" could enhance their "Pro-Environmental Attitudes".

Similarly, for increase in "*Environmental Awareness"*, the predicted "*Pro-Environmental Attitudes*" increased by approximately 0.289 or 28.9%. Confidence intervals indicated that we can be 95% certain that the slope to predict "*Environmental Awareness*" from "*Pro-Environmental Attitudes* "is between 0.289 and 0.539. The regression coefficients showed that for "*Pro-Environmental Attitudes*" increased by an average of 0.502 (50.2%) (β_1 = 0.441, SE = 0.064, t = 6.500, p < 0.001). This result indicated that "Pro-Environmental Attitudes" are a strongly significant predictor of "*Environmental Awareness*", supporting the hypothesis that increased "*Environmental Awareness*" is associated with "Pro-Environmental Attitudes". Overall, the findings suggest that encouraging local tourism community to engage in more "*Environmental Awareness*" could enhance their "Pro-Environmental Attitudes". In summary, the goal of multiple regression analysis is to identify which independent variables have the greatest influence on predicting the dependent variable. Thus, the results indicate that "environmental awareness" is the most significant factor influencing "pro-environmental" attitudes, accounting for 50.2%. This means that if researchers wish to change the behavior or attitudes of the local tourism community, they would gain more knowledge and awareness about the environments.

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.669[a]	(1) .448	(2) .446	.72736	.448	314.284	1	388	<.001	
2	.708[b]	(A) .502	(B) .499	.69153	.054	42.247	1	387	<.001	(7)+(G) 1.510

a. Predictors: (Constant), ENKX
b. Predictors: (Constant), ENKX, ENAX
c. Dependent Variable: PEAX

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	166.274	(11) 1	166.274	(3) 314.284	(4) <.001[b]
	Residual	205.274	(12) 388	.529		
	Total	371.548	389			
2	Regression	186.477	(K) 2	93.239	(C) 194.971	(D) <.001[c]
	Residual	185.070	(L) 387	.478		
	Total	371.548	389			

a. Dependent Variable: PEAX
b. Predictors: (Constant), ENKX
c. Predictors: (Constant), ENKX, ENAX

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	Upper Bound	Collinearity Statistics Tolerance	VIF
1	(Constant)	1.194	.122		9.759	<.001	.953	1.434		
	ENKX	.589	.033	.669	17.728	<.001	.524	.655	1.000	1.000
2	(Constant)	.891	.125		7.115	<.001	.645	1.137		
	ENKX	.260	.060	(13) .295	(5) 4.355	(6) <.001	(9) .143	(10) .377	.280	(8) 3.569
	ENAX	.414	.064	(M) .441	(E) 6.500	(F) <.001	(I) .289	(J) .539	.280	(H) 3.569

a. Dependent Variable: PEAX

REFERENCES

Anderson, D. R., Sweeney, D. J., Williams, T. A., Freeman, J., & Shoesmith, E. (2014). *Statistics for business and economics* (12th ed.). Cengage Learning, Inc.

Doane, D. P., & Seward, L. W. (2016). *Applied statistics in business and economics.* Mcgraw-Hill.

Groebner, D. F., Shannon, P. W., & Fry, P. C. (2018). *Business statistics: A decision-making approach.* Pearson.

Hair Jr, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate data analysis: A global perspective.* Prentice Hall and Pearson.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM).* Sage publications.

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity, 52*(4), 1957-1976. https://doi.org/10.1007/s11135-017-0584-6

GLOSSARY

- **Regression Equation:** An equation that expresses the linear relationship between independent variable and dependent variable.
- **Regression Slope Coefficient:** The average change in the dependent variable for a unit change in the independent variable. The slope coefficient may be positive, negative, or zero, depending on the relationship between the two variables.
- **Coefficient of Determination:** The portion of the total variation in the dependent variable that is explained by its relationship with the independent variable. The coefficient of determination is also called R-squared and is denoted as R^2.
- **R-square (R^2):** The squared correlation coefficient (R^2) is a very important statistic to explain the strength of the relationship we have between two variables.
- **Adjusted R-Squared:** A measure of the percentage of explained variation in the dependent variable in a multiple regression model that takes into account the relationship between the sample size and the number of independent variables in the regression model.
- **Simple Linear Regression:** The method of regression analysis in which a single independent variable is used to explain the variation in the dependent variable.
- **Multiple Regression:** extends simple regression to include several independent variables (called predictors).
- **Correlation Coefficient:** A visual display is a good first step in analysis, but we would also like to quantify the strength of the association between two variables. It is a quantitative measure of the strength of the linear relationship between two variables. The correlation ranges from -1.0 to +1.0. A correlation of {1.0 indicates a perfect linear relationship, whereas a correlation of 0 indicates no linear relationship.
- **Variance Inflation Factor—(VIF):** A measure of how much the variance of an estimated regression coefficient increases if the independent variables are correlated. A VIF equal to 1.0 for a given independent variable indicates that this independent variable is not correlated with the remaining independent variables in the model. The greater the multicollinearity, the larger the VIF.