

Final Project-2

Lecture 8

L8- Statistical Analysis- P4



Session Outline

- Hypotheses Testing
- Findings' Discussion

Hypotheses Testing

اختبار الفرضيات



SIMPLE AND MULTIPLE LINEAR REGRESSION

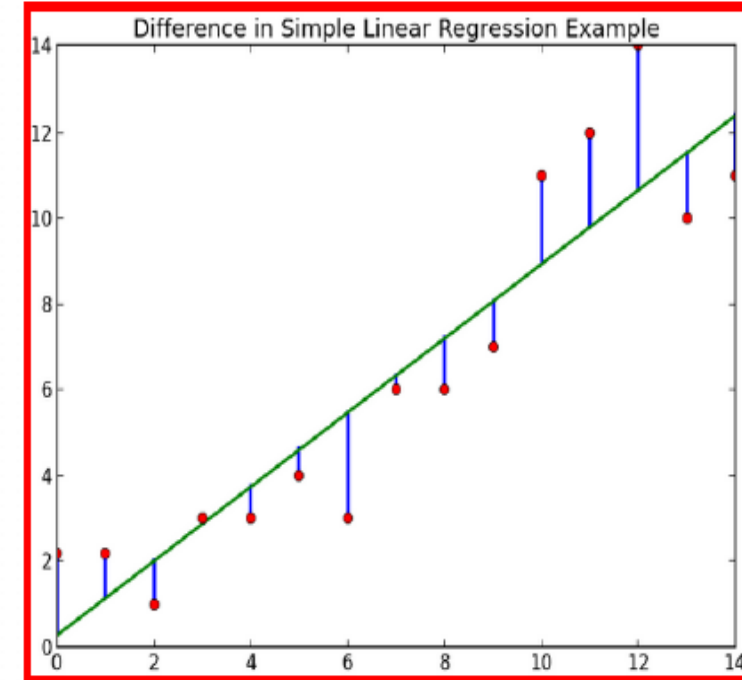
❑ INTRODUCTION TO REGRESSION ANALYSIS

❖ Regression analysis is used to:

- Predict the value of a dependent variable based on the value of at least one independent variable
- Explain the impact of changes in an independent variable on the dependent variable
- **Dependent variable:** the variable we wish to predict or explain. (Usually Y)
- **Independent variable:** the variable used to predictor explain the dependent variable. (Usually X)

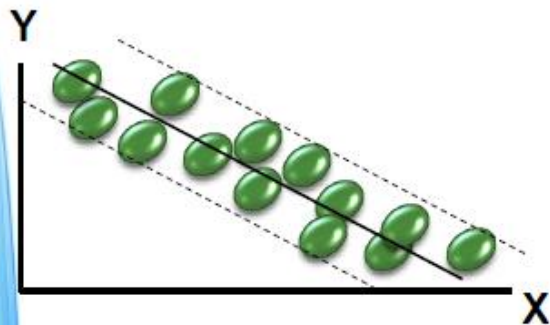
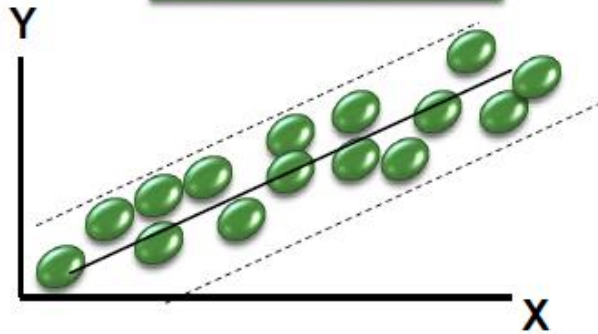
SIMPLE LINEAR REGRESSION MODEL

- ▶ Only **one independent variable**, X .
- ▶ Relationship between X and Y is described by a linear function.
- ▶ Changes in Y are assumed to be related to changes in X .

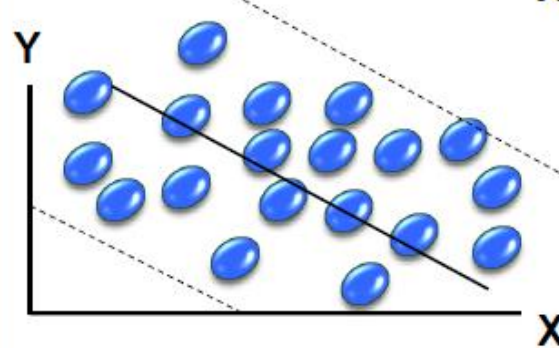
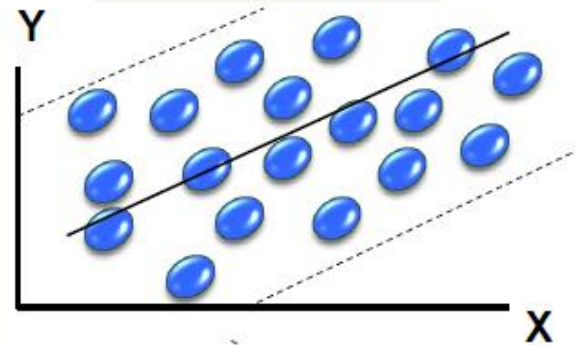


TYPES OF RELATIONSHIPS

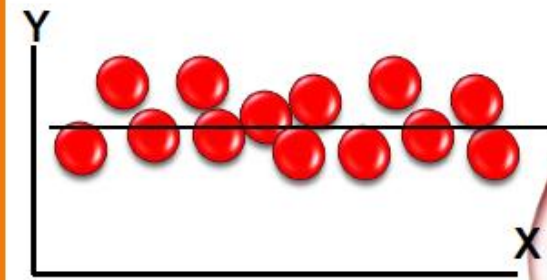
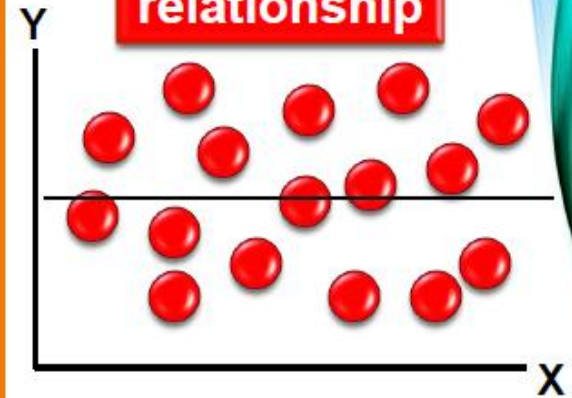
Strong relationships



Weak relationships



No relationship



SIMPLE LINEAR REGRESSION MODEL

The diagram illustrates the Simple Linear Regression Model equation, $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, with labels for its components:

- Dependent Variable**: Points to Y_i .
- Population Y intercept**: Points to β_0 .
- Population Slope Coefficient**: Points to β_1 .
- Independent Variable**: Points to X_i .
- Random Error term**: Points to ϵ_i .

The equation is displayed on a green background. A yellow bracket under $\beta_0 + \beta_1 X_i$ is labeled **Linear components**. A yellow bracket under ϵ_i is labeled **Random Error component**.

SIMPLE LINEAR REGRESSION EQUATION (PREDICTION LINE)

The simple linear regression equation provides an estimate of the population regression line

Estimated (or predicted) Y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

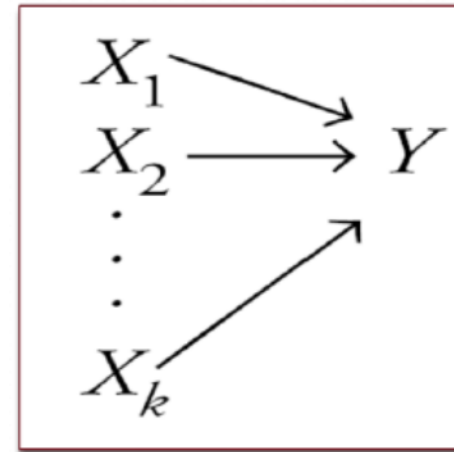
Value of X for observation i

$$\hat{Y}_i = b_0 + b_1 X_i$$

□ MULTIPLE LINEAR REGRESSION MODEL

Multiple regression simultaneously considers the influence of multiple explanatory variables on a response variable Y

- The intention is to look at the independent effect of each variable while "adjusting out" for other variables



THE MULTIPLE REGRESSION MODEL

Relationship between one dependent & 2 or more independent variables is a linear function

Population Y-intercept Population slopes Random error

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

Linear components Random Error component

Y is the dependent (Response) variable

X₁, X₂, ..., X_k are the independent (Explanatory) variables

□ R square estimate

- R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.
- R-squared is always between 0 and 100%
- 0% indicates that the model explains none of the variability of the response data around its mean.
- 100% indicates that the model explain all the variability of the response data around its mean.
- In general, the higher the R-squared, the better the model fits your data.

□ P value

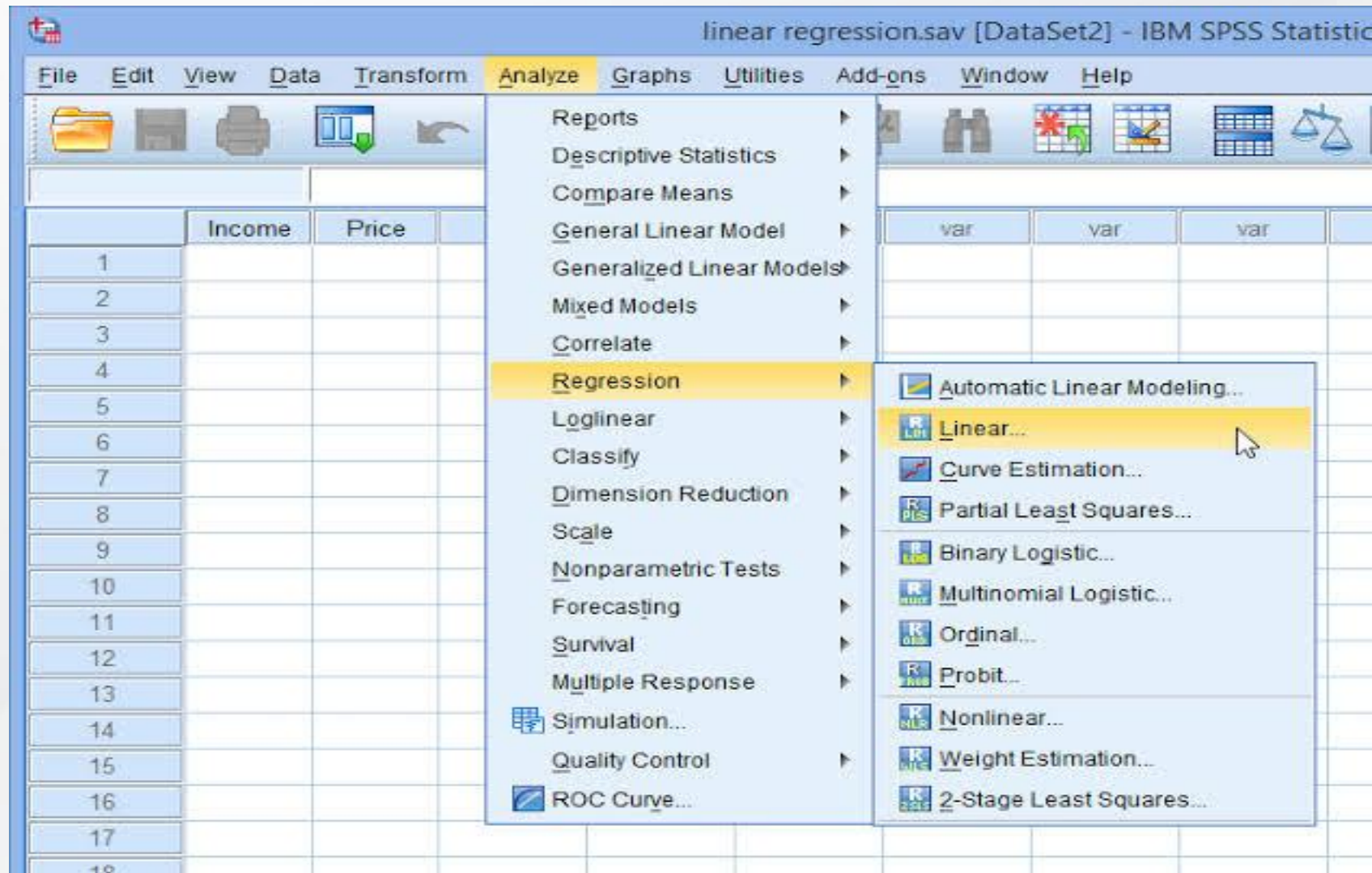
- The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect).
- A low p-value (<0.05) indicates that you can reject the null hypothesis. In other words, a predictor that has a low p-value is likely to be a meaningful addition to your model because changes in the predictor's value are related to changes in the response variable.
- Conversely, a larger (insignificant) p-value suggests that changes in the predictor are not associated with changes in the response.

□ Beta coefficient

- A standardized beta coefficient compares the strength of the effect of each individual independent variable to the dependent variable.
- The higher the absolute value of the beta coefficient, the stronger the effect.

SPSS

Regression in SPSS



Linear Regression

Income

Dependent: Price

Block 1 of 1

Previous Next

Independent(s): Income

Method: Enter

Selection Variable: Rule...

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options...

SPSS Statistics File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help Sat 7:50 AM David L

Multiple Linear Regression.sav [DataSet2] - IBM SPSS Statistics Data Editor

Output3 [Document3] - IBM SPSS Statistics Viewer

Output

- Log
- Regression
 - Time
 - Notes
 - Active Dataset
 - Variables Entered/Removed
 - Model Summary
 - ANOVA
 - Coefficients

[DataSet2] /Users/davidlongstreet/Documents/SPSS Examples/Multiple Linear Regression .sav

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Study_Time, IQ ¹		Enter

a. Dependent Variable: GPA
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Sd. Error of the Estimate
1	.954 ^a	.910	.890	.2654

a. Predictors: (Constant), Study_Time, IQ

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.389	2	3.194	45.353	.000 ^b
	Residual	.634	9	.070		
	Total	7.023	11			

a. Dependent Variable: GPA
b. Predictors: (Constant), Study_Time, IQ

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Sd. Error	Beta		
1					

IBM SPSS Statistics Processor is ready

Testing Mediation Effect via SPSS

There are two methods in testing indirect effect (mediation):

- 1– Baron and Kenny method
- 2- Hayes method (macro- process)

Baron & Kenny method

□ Baron and Kenny mediation steps

- ❖ The above steps of mediation is based on the four step mediation analysis test proposed by Baron and Kenny (1986), Judd and Kenny (1981), and James and Brett (1984).
- ❖ Thus, to indicate mediation four steps are to be analyzed-

Step 1: Show that the causal variable is correlated with the outcome. Use Y as the criterion variable in a regression equation and X as a predictor (estimate and test path c in the above figure). This step establishes that there is an effect that may be mediated.

Step 2: Show that the causal variable is correlated with the mediator. Use Mas the criterion variable in the regression equation and X as a predictor (estimate and test path a). This step essentially involves treating the mediator as if it were an outcome variable.

□ Baron and Kenny mediation steps

Step 3: Show that the mediator affects the outcome variable. Use Y as the criterion variable in a regression equation and X and M as predictors (estimate and test path b). It is not sufficient just to correlate the mediator with the outcome because the mediator and the outcome may be correlated because they are both caused by the causal variable X. Thus, the causal variable must be controlled in establishing the effect of the mediator on the outcome.

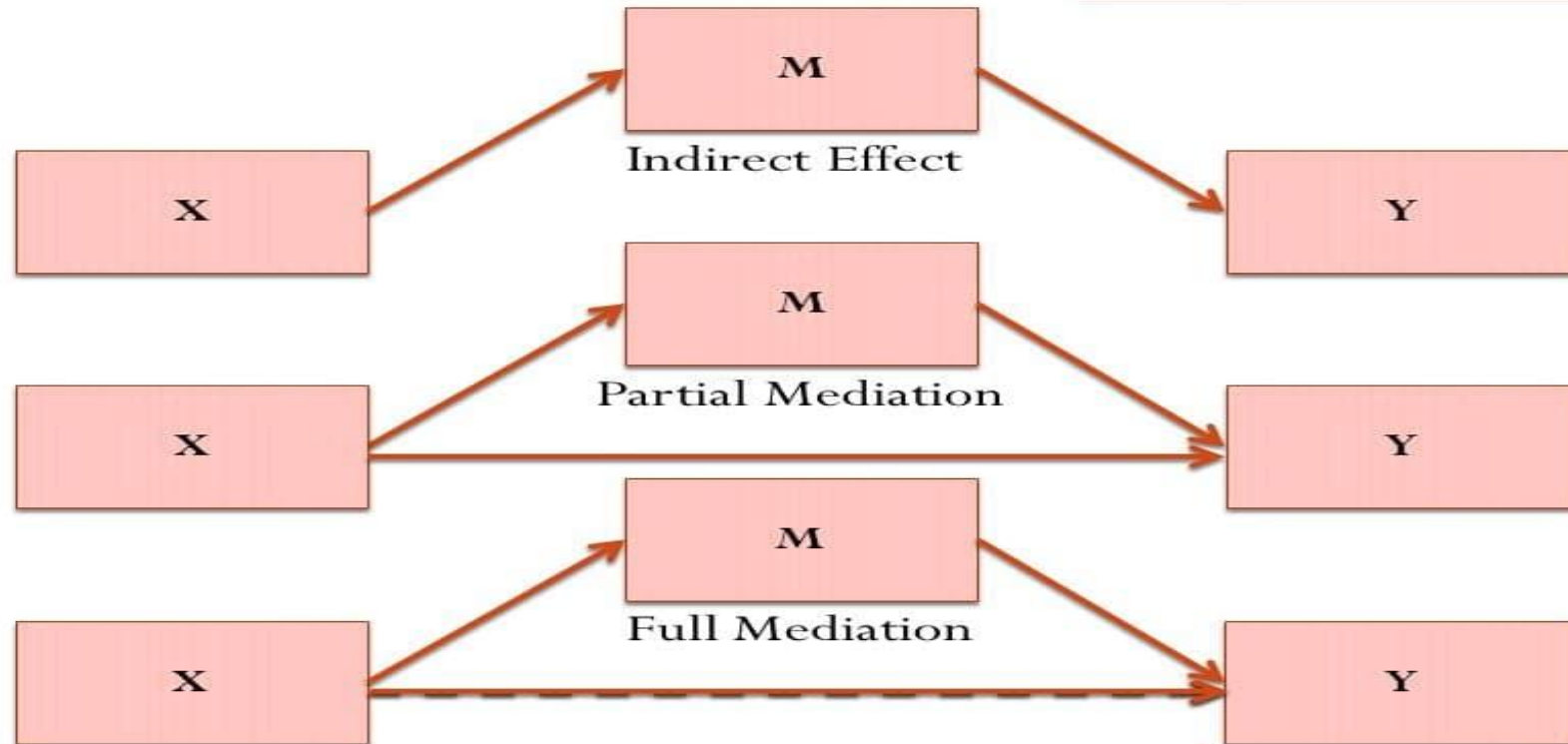
Step 4: To establish that M completely mediates the X-Y relationship, the effect of X on Y controlling for M (path c) should be zero (see discussion below on significance testing). The effects in both Steps 3 and 4 are estimated in the same equation.

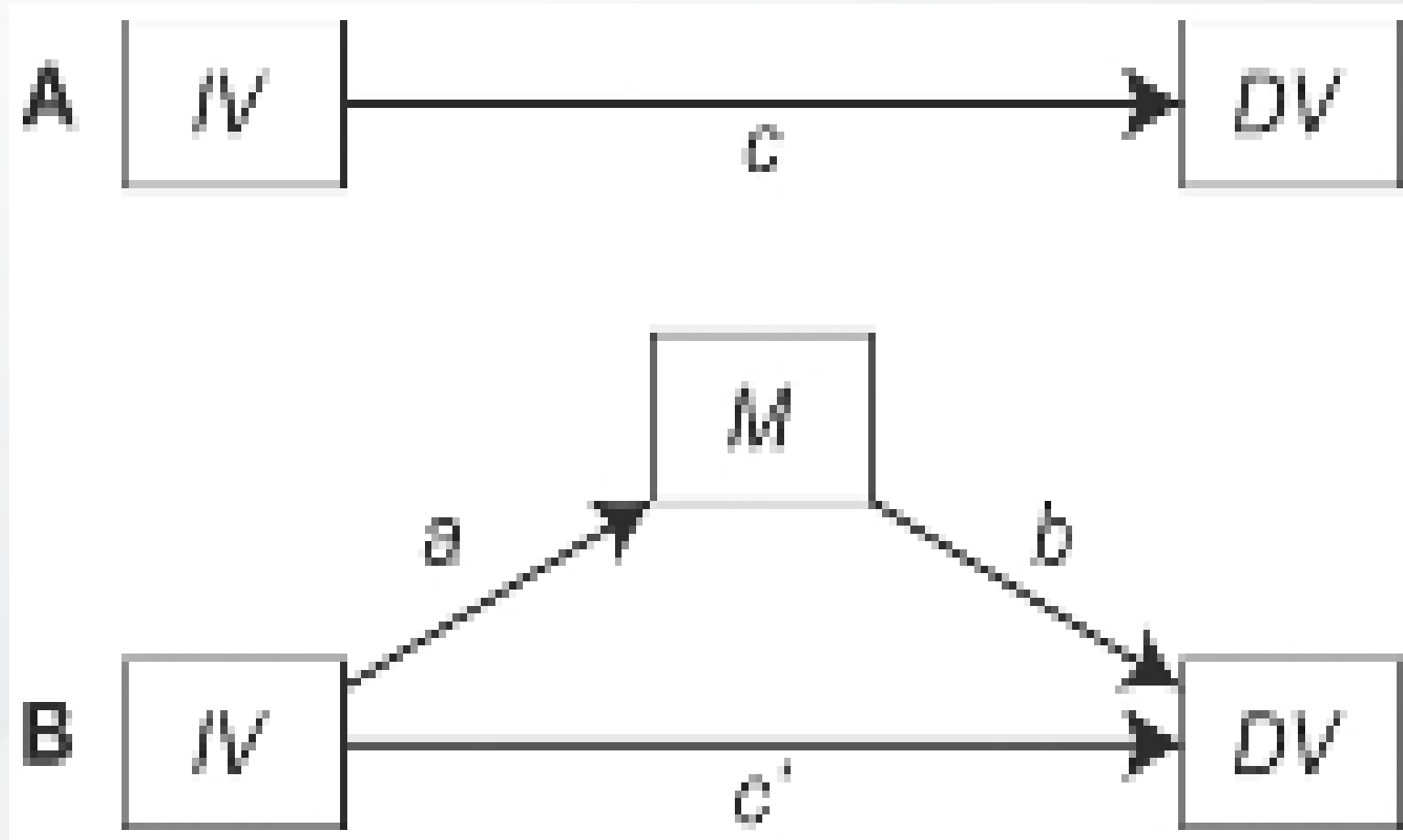
□ Baron and Kenny mediation steps

❖ Final mediation decision

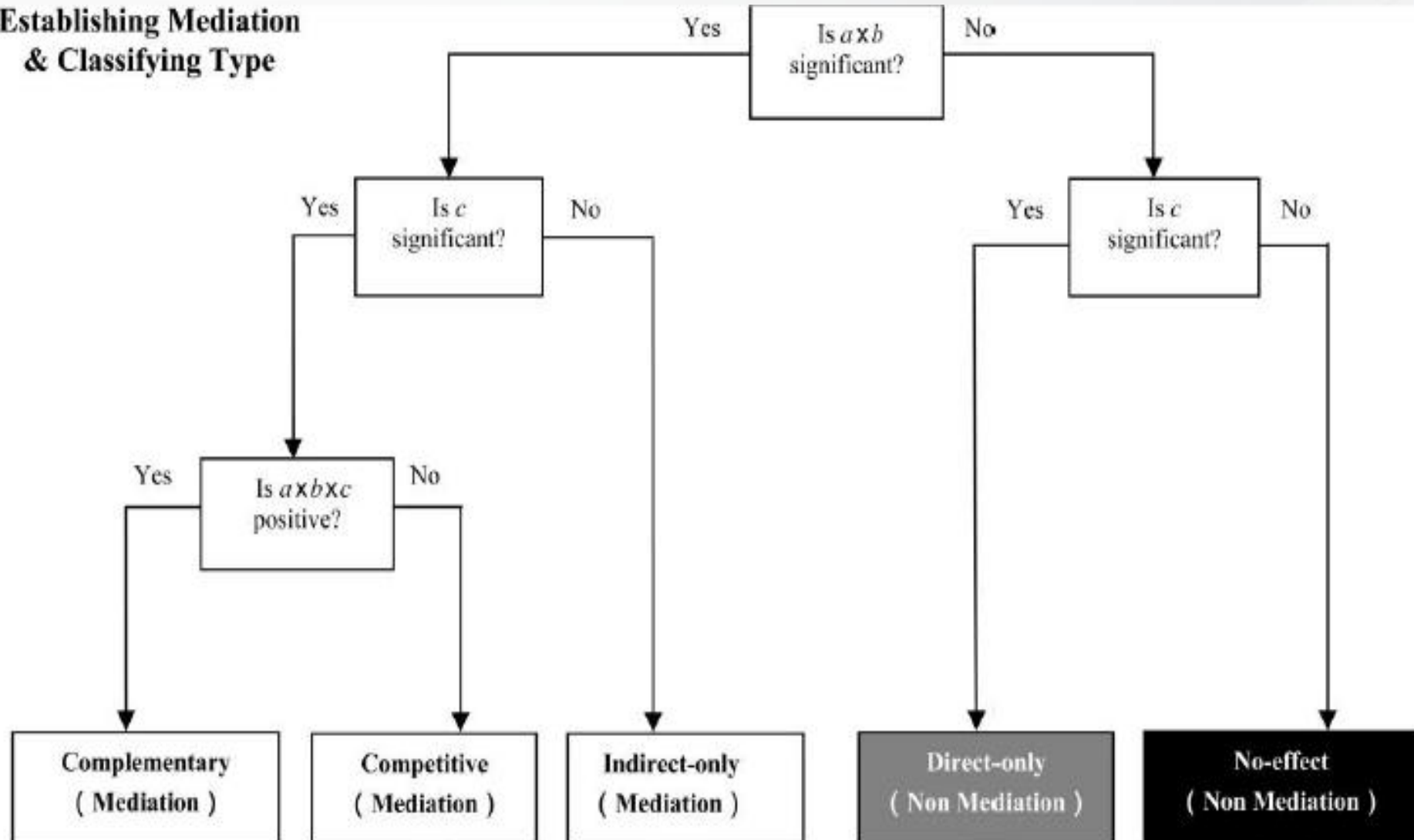
- If all four of these steps are met, then the data are consistent with the hypothesis that variable M completely mediates the X-Y relationship
- If the first three steps are met but the Step 4 is not, then partial mediation is indicated.
- Meeting these steps does not, however, conclusively establish that mediation has occurred because there are other (perhaps less plausible) models that are consistent with the data.
- Some of these models are considered later in the Specification Error section.

Types of Mediation





Establishing Mediation & Classifying Type



□ References:

- Hayes, A.F., 2009. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs*, 76(4), pp.408-420.
- Shrout, P.E. and Bolger, N., 2002. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods*, 7(4), p.422.
- Zhao, X., Lynch Jr, J.G. and Chen, Q., (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of consumer research*, 37(2), pp.197-206.

View Data Transform **Analyze** Direct Marketing Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...

Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
PROCESS v3.5 by Andrew F. Hayes
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling (CATREG)...

	A1	A2	M2	M3	M4	M5	O1	O2	O3	O4
	4		4	3	4	2	3	3	4	3
	4		3	2	4	3	3	3	4	4
	3		2	2	4	2	3	3	4	3
	1		1	1	4	1	1	1	1	2
	2					2	1	2	2	2
	3					2	2	4	4	4
	2					1	1	1	1	1
	2					1	1	1	3	2
	5					4	4	4	2	2
	4					1	3	3	4	3
	5					3	2	2	2	1
	5					3	3	3	4	4
	5					5	5	2	2	2
	4					2	3	2	3	2
	3					2	2	3	4	4
	2					1	1	2	2	2
	4					2	1	3	5	3
	4					2	4	4	5	3
	1					2	2	3	4	4
	4					3	3	4	3	3
						5	5	5	4	4

A3	M1	M2	M3	M4	M5	O1	O2	O3	O4	SQ1
3	3								4	3
3	4								4	4
3	2								4	3
2	1								1	2
2	2								2	2
4	4								4	4
4	1								1	1
4	2								3	2
5	4								2	2
2	3								4	3
5	4								2	1
4	3								4	4
5	5								2	2
4	5								3	2
3	4								4	4
2	2								2	2
1	4								5	3
4	4								5	3
3	3								4	4
4	3								3	3
									4	4

Linear Regression

SVQ4

SVQ5

OA1

OA2

OA3

OA4

OA5

OA6

OA7

OA8

Ability

Motivation

Opportunity

SQ

IQ

SVQ

KMS

HPWS

Dependent:

OA

Block 1 of 1

Previous

Next

Independent(s):

HPWS

KMS

Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

OK

Paste

Reset

Cancel

Help

Statistics...

Plots...

Save...

Options...

Bootstrap...

Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.665 ^a	.442	.404	.698

a. Predictors: (Constant), KMS, HPWS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11.561	2	5.780	11.862	.000 ^b
	Residual	14.619	30	.487		
	Total	26.180	32			

a. Dependent Variable: OA

b. Predictors: (Constant), KMS, HPWS

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.115	.709		2.984	.006
	HPWS	.188	.162	.180	1.158	.256
	KMS	.823	.228	.560	3.610	.001

a. Dependent Variable: OA

Hayes Method Macro- PROCESS

- Macro- PROCESS method is originated by Hayes, (2012), as it is most flexible and effective tool in testing the indirect effect (the mediation effect) (MacKinnon, Fairchild, and Fritz, 2007; Hayes, 2009).
- The advantage with using PROCESS tests that the mediation effect is assessed with the aid of bootstrapping.
- Indirect effect to be tested, bootstrapping steps (n= 5000 resample) and confidence interval at 95%.
- The mediation effect can be considered statistically significant only when (zero) doesn't exist between their range of confidence intervals, namely: BootLLCI "Bootstrap Lower Limit Confidence Interval" and BootULCI "Bootstrap Upper Limit Confidence Interval" (Hayes, 2012; Vinod, et al., 2015).

- **In Macro- PROCESS, select the number of the model as follows:**
 - ✓ Model 4 for mediation
 - ✓ Model 1 for moderation
 - ✓ Model 5 for mediation/ moderation

View Data Transform **Analyze** Direct Marketing Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control

M2 M3 M4 M5 O1 O2 O3 O4

4								
4								
3								
1								
2								
3								
2								
2								
5								
4								
5								
5								
5								
4								
3								
2								
4								
4								
1								

Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
PROCESS v3.5 by Andrew F. Hayes
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling (CATREG)...

PROCESS_v3.5

	A3	M
4	3	
4	3	
2	3	
3	2	
2	2	
4	4	
3	4	
3	4	
5	5	
3	2	
4	5	
5	4	
5	5	
5	4	
2	3	
2	2	
2	1	
4	4	
4	3	
4	4	
5	4	
4	4	

Variables:

- A1
- A2
- A3
- M1
- M2
- M3
- M4
- M5
- O1
- O2
- O3
- O4
- SQ1
- SQ2

Model number: 4

Confidence intervals: 95

Number of bootstrap samples: 5000

☐ Save bootstrap estimates

☐ Bootstrap inference for model coefficients

Y variable: OA

X variable: HPWS

Mediator(s) M: KMS

Covariate(s):

Moderator variable W:

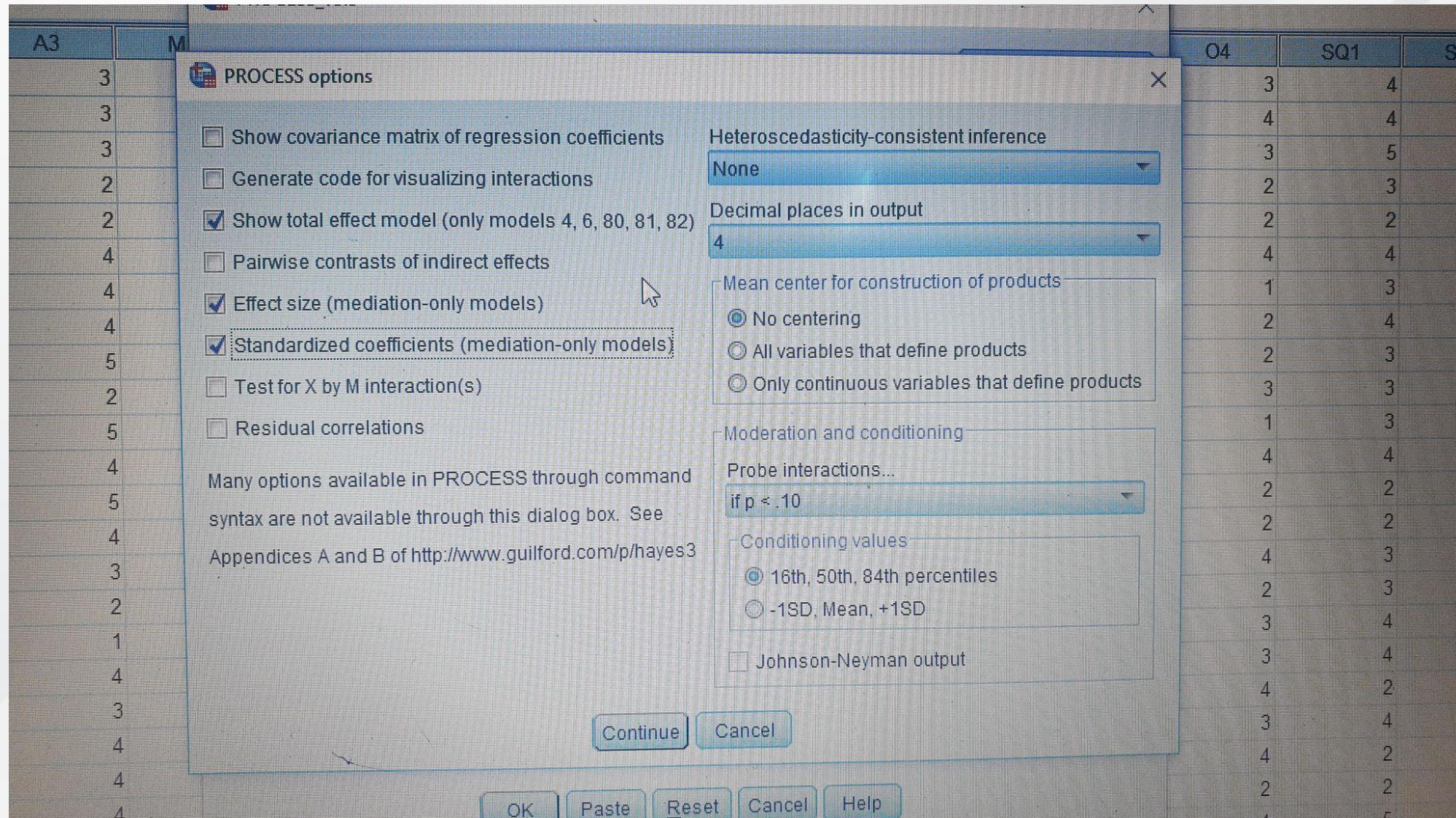
Moderator variable Z:

Do not use PASTE button

About
Options
Multicategorical
Long variable names

	O4	SQ1
	3	4
	4	4
	3	5
	2	3
	2	2
	4	4
	1	3
	2	4
	2	3
	3	3
	1	3
	4	4
	2	2
	2	2
	4	3
	2	3
	3	4
	3	4
	4	2
	3	4
	4	2
	2	2

OK Paste Reset Cancel Help



Matrix

Title

Notes

Active Dataset

Text Output

Log

Matrix

Title

Notes

Active Dataset

Text Output

KMS

Model Summary

R

R-sq

MSE

F

df1

df2

P

.4760

.2265

.3027

9.0796

1.0000

31.0000

.0051

Model

coeff

se

t

p

LLCI

ULCI

constant

2.2740

.3810

5.9679

.0000

1.4968

3.0511

HPWS

.3390

.1125

3.0132

.0051

.1096

.5685

Standardized coefficients

coeff

HPWS .4760

OUTCOME VARIABLE:

OA

Model Summary

R

R-sq

MSE

F

df1

df2

P

.6645

.4416

.4873

11.8616

2.0000

30.0000

.0002

Model

coeff

se

t

p

LLCI

ULCI

constant

2.1145

.7087

2.9837

.0056

.6672

3.5619

HPWS

.1880

.1623

1.1580

.2560

-.1435

.5195

KMS

.8225

.2279

3.6095

.0011

.3571

1.2879

OUTCOME VARIABLE:							
OA							
Model Summary							
	R	R-sq	MSE	F	df1	df2	p
	.4462	.1991	.6764	7.7047	1.0000	31.0000	.0093
Model							
	coeff	se	t	p	LLCI	ULCI	
constant	3.9849	.5696	6.9964	.0000	2.8233	5.1466	
HPWS	.4668	.1682	2.7757	.0093	.1238	.8099	
Standardized coefficients							
	coeff						
HPWS	.4462						
***** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y *****							
Total effect of X on Y							
Effect	se	t	p	LLCI	ULCI	c_ps	c_cs
.4668	.1682	2.7757	.0093	.1238	.8099	.5161	.4462
Direct effect of X on Y							
Effect	se	t	p	LLCI	ULCI	c'_ps	c'_cs
.1880	.1623	1.1580	.2560	-.1435	.5195	.2078	.1796
Indirect effect(s) of X on Y:							
Effect	BootSE	BootLLCI	BootULCI				
KMS	.2789	.1012	.0966	.4982			

Moderation Effect via SPSS

□ Terminology

- Criterion Variable (dependant variable)
- Predictor Variable (Independent variable)

$$y = a + b_1x$$

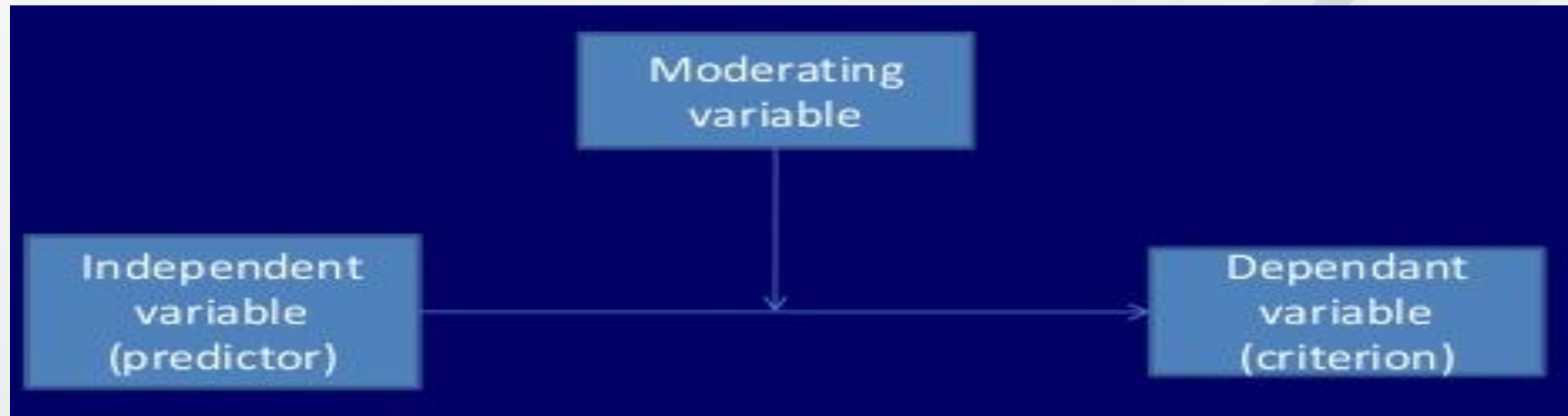
y = criterion variable x = predictor variable

□ Moderation

- A third variable changes the relationship between two related variable

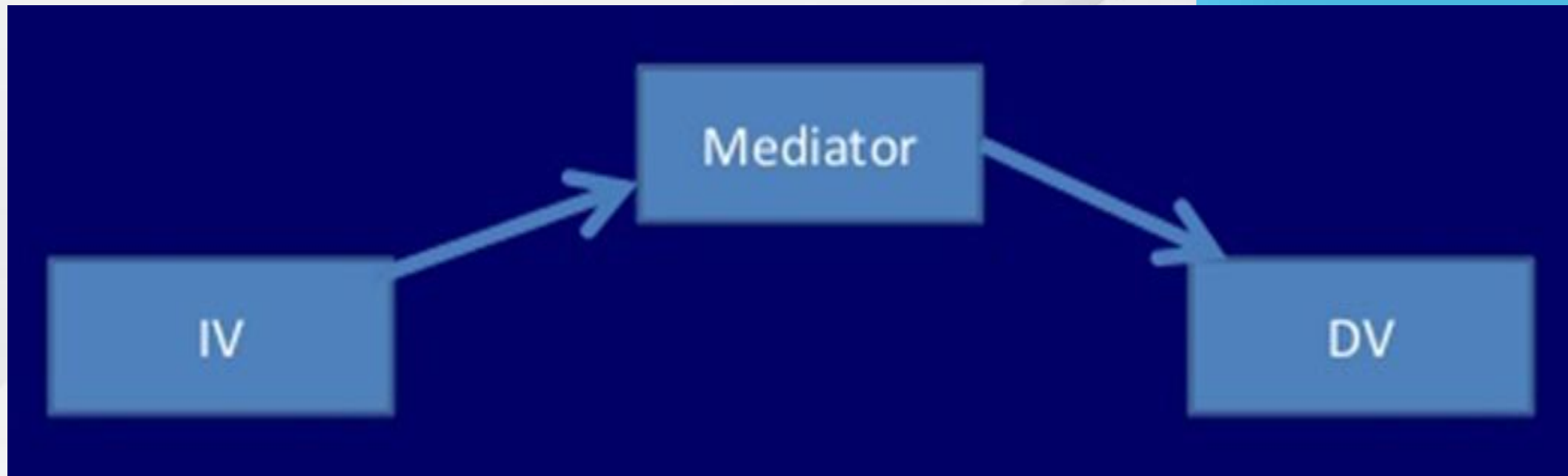
In other words

- Relationship between two variables changes with the level of another variable of construct



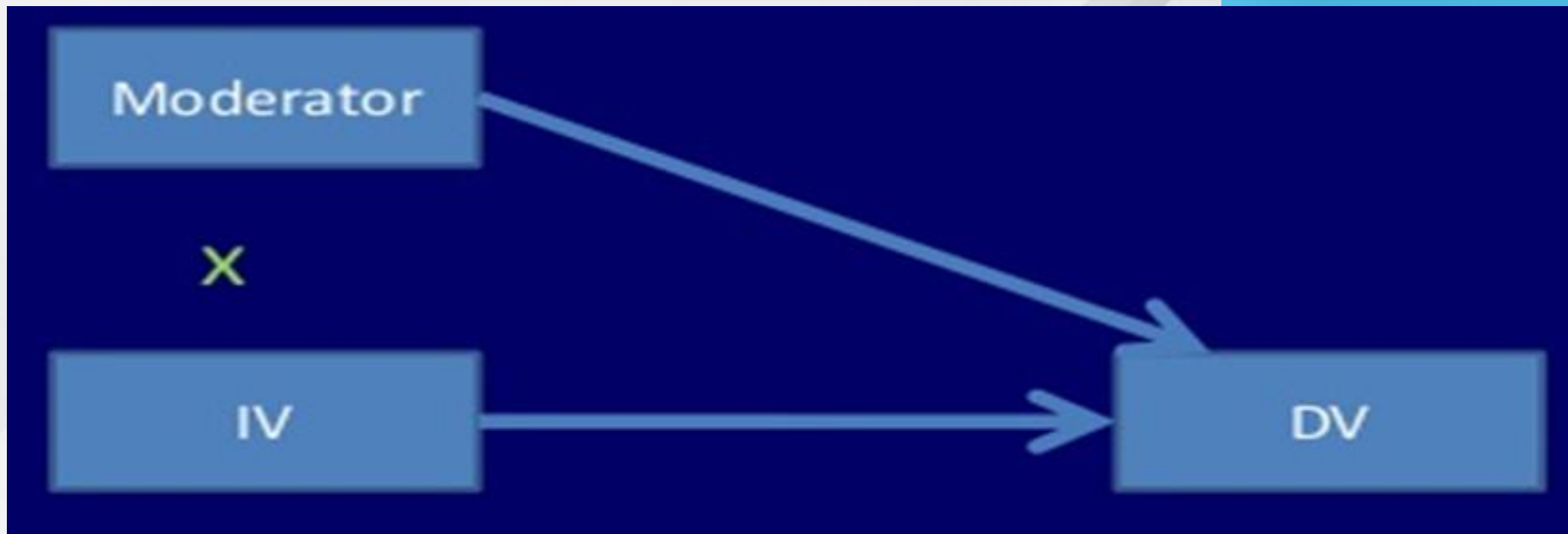
❑ Moderator v/s Mediation

- Mediator variable is consequent of predictor variable.



❑ Moderator v/s Mediation

- Moderator variable is at same level as predictor variable.



□ Aiken and West Approach

- **Step1:** Regress on dependent variable the independent variable and moderator variable
- **Step2:** Center all the continuous variables
- **Step3:** Multiply the independent and moderator variable and create interaction variable
- **Step 4:** Regress on dependent variable the independent variable, moderator variable and Interaction variable

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Sage.

□ Aiken and West Approach

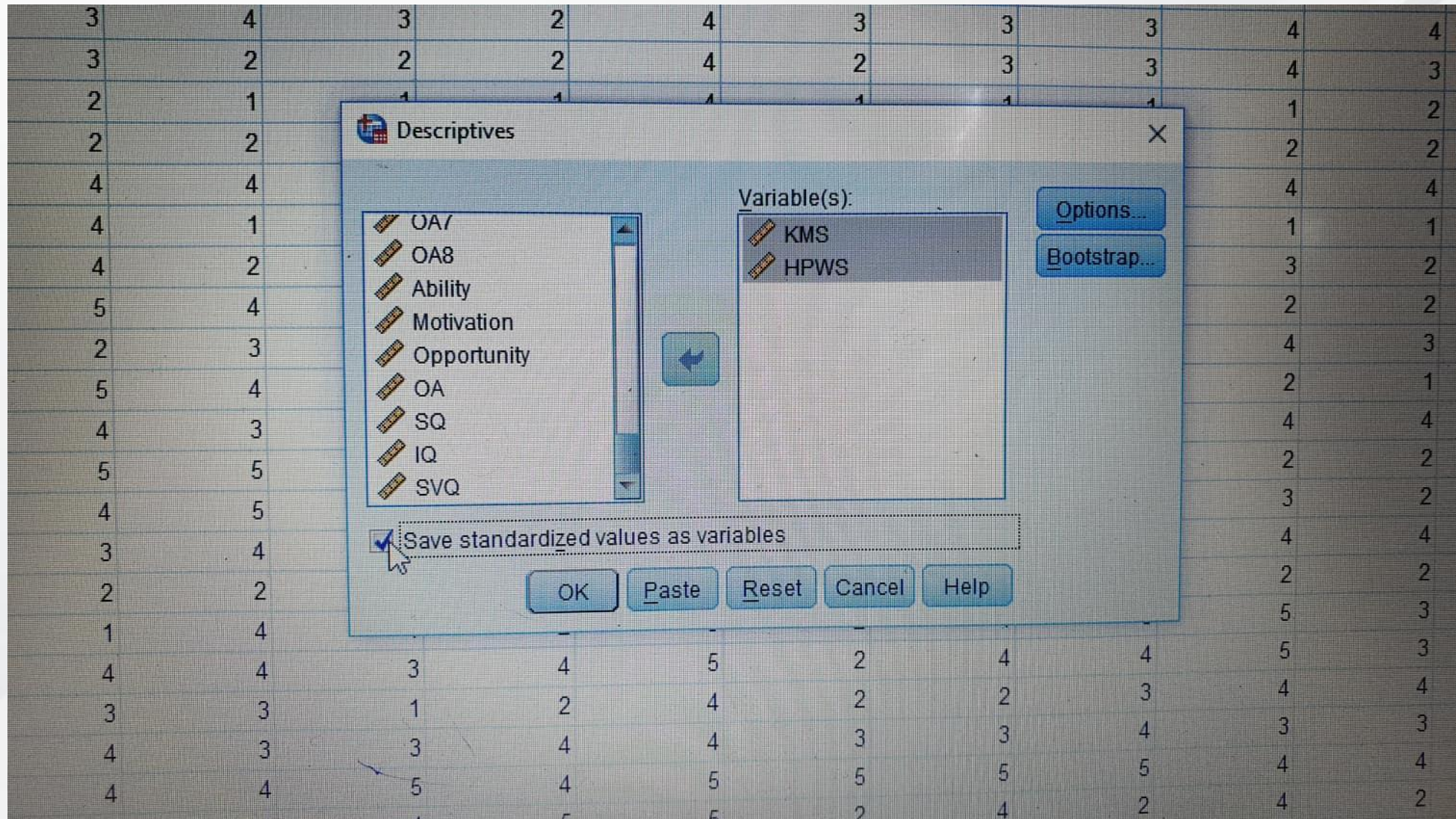
- **Step 5:** Check whether interaction variable is significant else no Moderation
- **Step 6:** ΔR^2
- **Step 7:** If moderator is continuous variable classify into meaning full groups or classify the same based at $+ \geq 1 \sigma$ level
- **Step 8:** Plot Graphs and interpret the same with significant slopes

Independent Variable	Moderation Variable	Interaction Variable	Moderation Effect
Significant	Significant	Significant	Yes
Significant	Significant	Not Significant	No
Significant	Not Significant	Not Significant	No
Not Significant	Significant	Significant	Yes

iv [Dataset1] - IBM SPSS Statistics Data Editor

Analyze | Descriptive Statistics | **Descriptives...**

	A1	A2	M4	M5	O1	O2	O3	O4
	4		4	2	3	3	4	
	4		4	3	3	3	4	
	3		4	2	3	3	4	
	1		4	1	1	1	1	
	2		2	2	1	2	2	
	3		4	2	2	4	4	
	2		1	1	1	1	1	
	2		3	1	1	1	3	
	5		4	4	4	4	2	
	4		4	3	4	1	3	
	5		4	3	3	2	2	
	4		4	4	3	3	4	
	5		3	3	3	3	2	
	5		5	5	5	5	2	
	5		4	5	4	2	3	
	4		4	4	2	2	3	
	3		4	3	1	1	2	
	2		3	2	1	1	3	
	4		4	5	2	4	4	
			3	4	2	2	3	



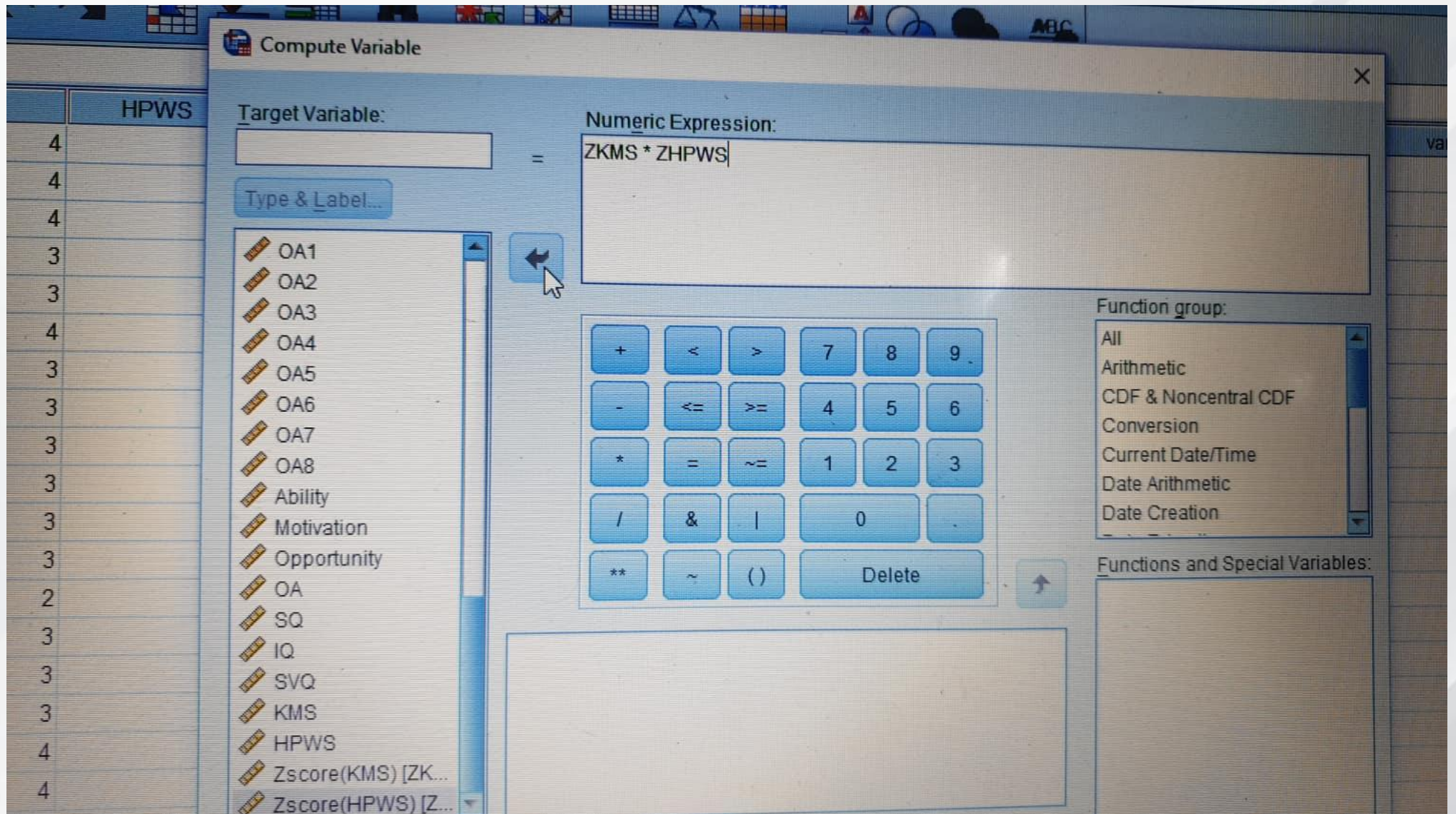
KMS	HPWS	ZKMS	ZHPWS	var	var	var	var	var
4	3	.99831	.06427					
4	3	.99831	.16067					
4	3	1.77164	-.61055					
3	2	-.70303	-1.96017					
3	2	-.70303	-1.67097					
4	4	.99831	.35347					
3	2	-.93503	-1.57456					
3	2	-.54837	-1.38176					
3	4	-.62570	.73908					
3	3	-.62570	-.22494					
3	3	-.70303	-.03213					
3	4	.07030	.35347					
2	4	-1.78570	1.12469					
3	4	-.70303	.35347					
3	3	-.93503	-.03213					
3	2	-.70303	-1.47816					
4	3	.30230	-.32134					
4	4	1.30764	.64268					
3	3	-.70303	-.61055					

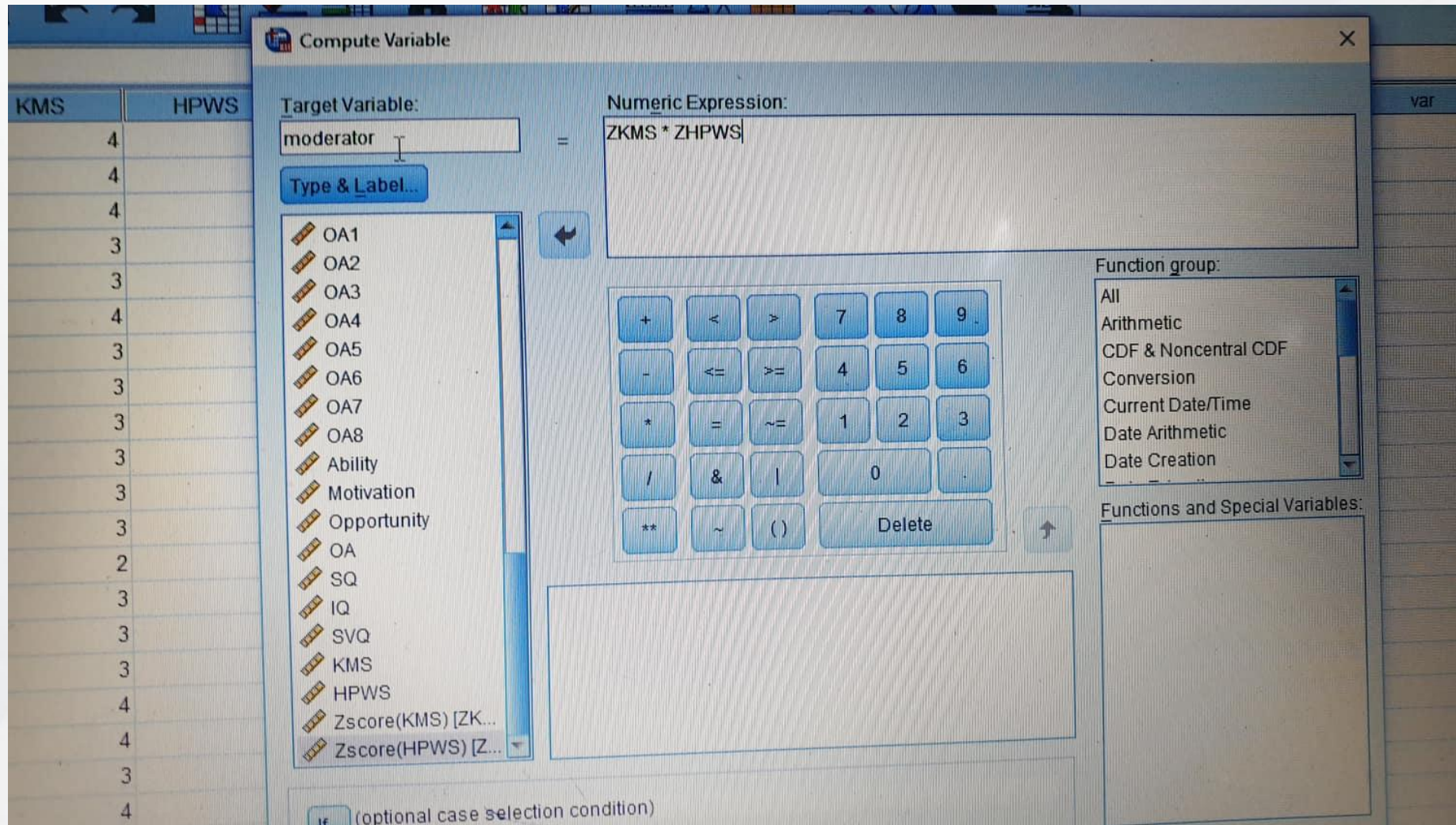
*SPSS Original.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

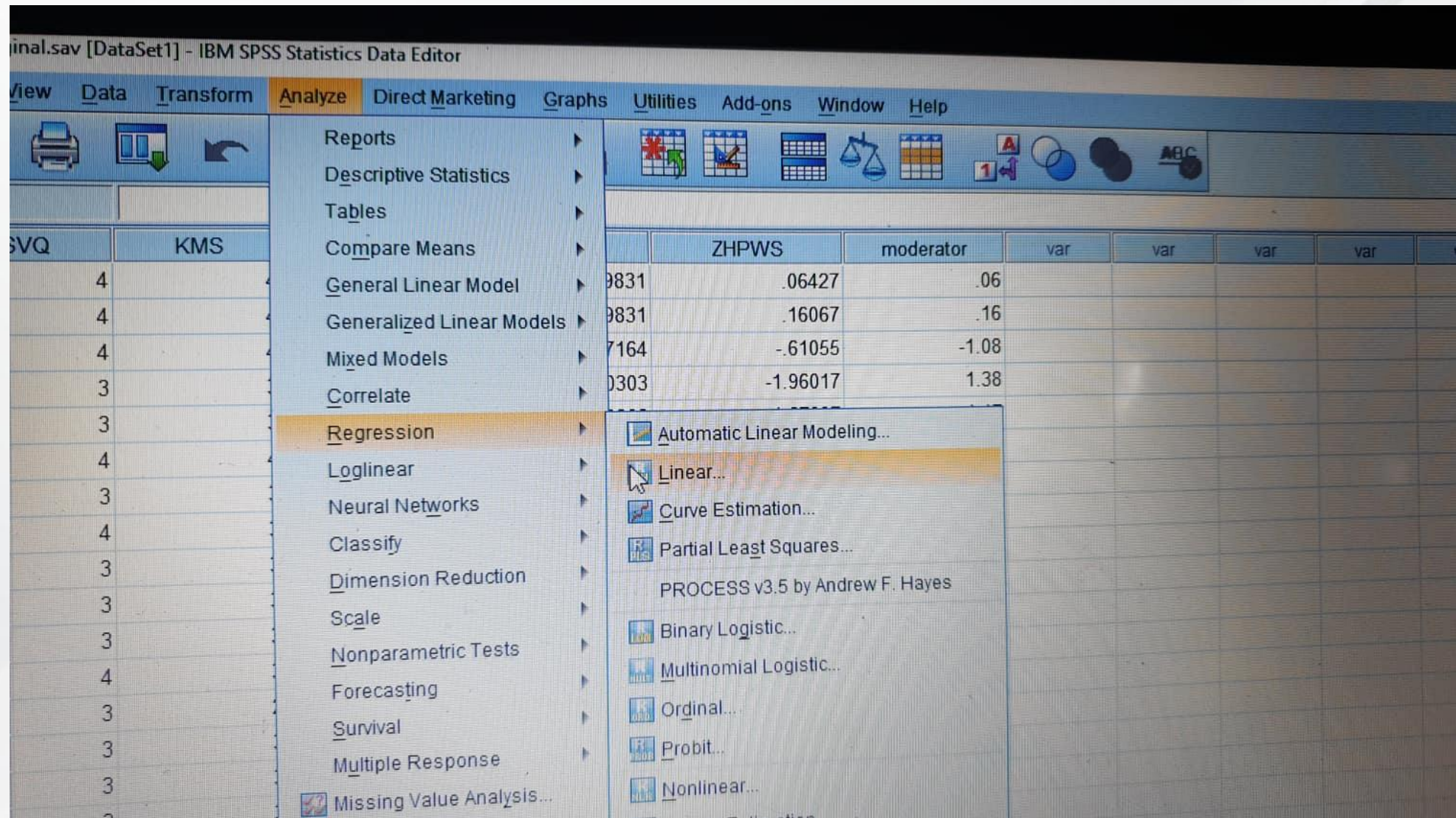
Compute Variable...
Count Values within Cases...
Shift Values...
Recode into Same Variables...
Recode into Different Variables...
Automatic Recode...
Visual Binning...
Optimal Binning...
Prepare Data for Modeling
Rank Cases...
Date and Time Wizard...
Create Time Series...
Replace Missing Values...
Random Number Generators...
Run Pending Transforms Ctrl+G

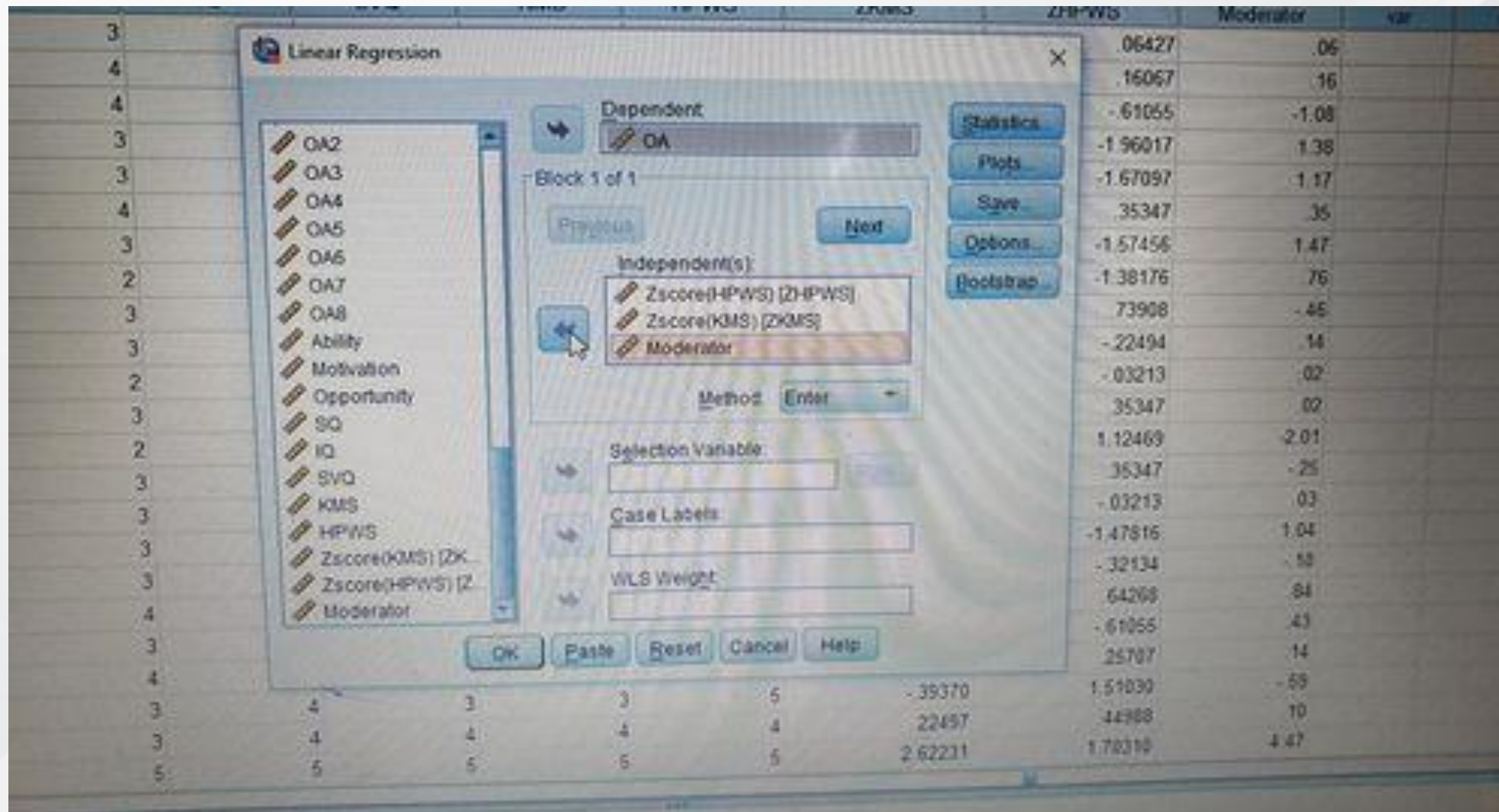
	SVQ		ZHPWS	var	var	v
1	4	831	.06427			
2	4	831	.16067			
3	4	164	-.61055			
4	3	303	-1.96017			
5	3	303	-1.67097			
6	4	831	.35347			
7	3	503	-1.57456			
8	4	837	-1.38176			
9	3	570	.73908			
10	3	570	-.22494			
11	3	303	-.03213			
12	4	030	.35347			
13	3	570	1.12469			





	KMS	HPWS	ZKMS	ZHPWS	moderator	var	var	var	va
4	4	3	.99831	.06427	.06				
4	4	3	.99831	.16067	.16				
4	4	3	1.77164	-.61055	-1.08				
3	3	2	-.70303	-1.96017	1.38				
3	3	2	-.70303	-1.67097	1.17				
4	4	4	.99831	.35347	.35				
3	3	2	-.93503	-1.57456	1.47				
4	3	2	-.54837	-1.38176	.76				
3	3	4	-.62570	.73908	-.46				
3	3	3	-.62570	-.22494	.14				
3	3	3	-.70303	-.03213	.02				
4	3	4	.07030	.35347	.02				
3	2	4	-1.78570	1.12469	-2.01				
3	3	4	-.70303	.35347	-.25				
3	3	3	-.93503	-.03213	.03				
3	3	2	-.70303	-1.47816	1.04				
4	4	3	.30230	-.32134	-.10				
4	4	4	1.30764	.64268	.84				
3	3	3	-.70303	-.61055	.43				
3	4	4	.53431	.25707	.14				





set
ntered
mary

(KMS)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11.771	3	3.924	7.898	.001 ^b
	Residual	14.408	29	.497		
	Total	26.180	32			

a. Dependent Variable: OA

b. Predictors: (Constant), Moderator, Zscore(HPWS), Zscore(KMS)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.554	.136		40.782	.000
	Zscore(HPWS)	.130 ^{**}	.150	.144	.867	.393
	Zscore(KMS)	.550	.156	.608	3.514	.001
	Moderator	-.083	.128	-.100	-.652	.520

a. Dependent Variable: OA

Findings` Discussion



Understanding the Research Discussion

- ❑ The research discussion chapter is a critical component of a thesis, where the findings are analyzed and interpreted in the context of existing literature.
- ❑ It provides a platform to demonstrate the significance of the research and its implications for the field.

Structuring Your Discussion

- ❑ Organize your discussion chapter by presenting the research findings, analyzing the results, and discussing their implications.
- ❑ A clear and logical structure enhances the readability and coherence of your thesis

Analyzing Research Findings

- ❑ Utilize statistical analysis and qualitative interpretation to analyze the research findings.
- ❑ This in-depth analysis provides a comprehensive understanding of the results and their significance within the research context

Interpreting the Results

- ❑ Interpret the research results within the framework of existing literature and theoretical perspectives.
- ❑ This critical analysis demonstrates a deep understanding of the research outcomes and their implications

Contextualizing the Findings

- ❑ Place your research findings within the broader context of the field, highlighting their relevance and potential impact.
- ❑ This contextualization strengthens the significance of the research within the academic community.

Engaging with Literature

- ❑ Engage with relevant literature to support and contextualize the research findings within the broader academic discourse.
- ❑ A comprehensive literature review strengthens the scholarly foundation of the discussion chapter.



Thank You