LEARNING OBJECTIVES

Upon completing part B of this session, you should be able to do the following:

- Incorporate qualitative variables into the regression model by using dummy variables
- Use variable transformations to model nonlinear relationships
- Test and interpret interaction effects

QUALITATIVE (DUMMY) VARIABLES

- In many situations we must work with qualitative independent variables such as gender (male, female), method of payment (cash, cheque, credit card), etc.
- For example, X_2 might represent gender where $X_2 = 0$ indicates male and $X_2 = 1$ indicates female.
- A variable such as this is called a **dummy** variable. A dummy variable has two outcomes.

DUMMY-VARIABLE MODEL EXAMPLE (WITH 2 LEVELS)

- As an extension of the problem involving the management staff salary survey, suppose that they also believes that the annual salary is related to whether the individual has a graduate degree.
- The years of experience, the score on the aptitude test, whether the individual has a relevant graduate degree, and the annual salary for each of the sampled 20 management staff are shown on the next slide.

BLITZ MANAGEMENT SALARY SURVEY
SAMPLE DATA

Exper.	Score	Degr.	Salary	Exper.	Score	Degr.	Salary
4	78	No	76.8	9	88	Yes	121.6
7	100	Yes	137.6	2	73	No	85.12
1	86	No	75.84	10	75	Yes	115.84
5	82	Yes	109.76	5	81	No	101.12
8	86	Yes	114.56	6	74	No	92.8
10	84	Yes	121.6	8	87	Yes	108.8
0	75	No	71.04	4	79	No	96.32
1	80	No	73.92	6	94	Yes	108.48
6	83	No	96	3	70	No	90.24
6	91	Yes	105.6	3	89	No	96

BLITZ MANAGEMENT SALARY SURVEY
DUMMY-VARIABLE MODEL

The new variable (Degree) added holds categorical data.

To include it into a regression model we need to set it up as a ’dummy variable’.

X_3 = 0 if individual does not have a graduate degree.

1 if individual does have a graduate degree.

BLITZ MANAGEMENT SALARY SURVEY
REGRESSION OUTPUT

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	25.424	23.619	1.076	0.298	-24.646	75.493
Experience (Years)	3.672	0.952	3.856	0.001	1.653	5.691
Aptitude Score	0.630	0.288	2.191	0.044	0.020	1.240
Degree(Num)	$\boxed{7.297}$	6.357	1.148	0.268	-6.179	20.774

$b_3 = 7.297$

Interpretation:

All other factors being equal, a management staff with a degree will earn, on average, $7,297 more than one without.

Note: The *p*-value = 0.268 > 0.05 and so there is not enough evidence to conclude that $\beta_3 \neq 0$.

BLITZ MANAGEMENT SALARY SURVEY
INDIVIDUAL T-TEST

- As the variable (Degree) is not significant we would *usually* decide to exclude it from the model.
- However, if common sense or theory tells us to leave a variable in a model, we might do so, even though the p-value might be over 10% or over 20% or even higher.

A BRIEF REVIEW:

STEPS IN MODEL BUILDING

1. Scatter diagrams to identify potential independent (*x)* variables.
2. Correlation analysis and check for multi-collinearity.
3. Create regression model with chosen independent (*x*) variables.
4. Check for significance of model and for each variable.
5. If necessary add/subtract variable from model.
6. Perform residual analysis.
7. Use model.

MODEL BUILDING

STEPWISE REGRESSION

- **Idea:**

  Develop the least squares regression equation in steps, either through forward selection, backward elimination, or through standard stepwise regression.

- **Logic:**

  The coefficient of partial determination is the measure of the marginal contribution of each independent variable, given that other independent variables are in the model.

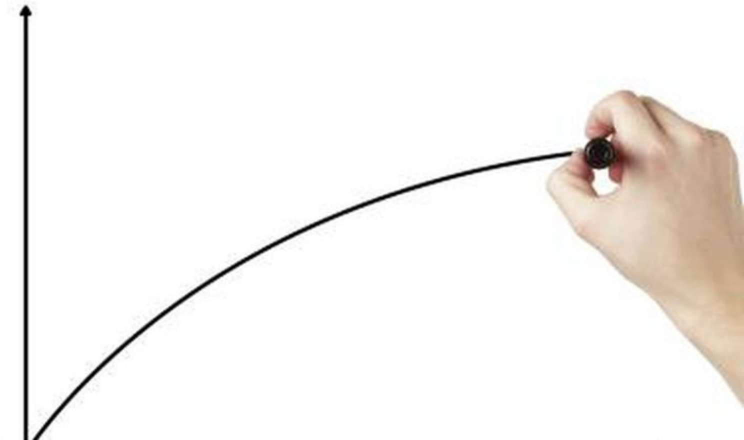
NONLINEAR RELATIONSHIPS

- The relationship between the dependent variable and an independent variable may not be linear
- Useful when scatter diagram indicates non-linear relationship.
- Example:

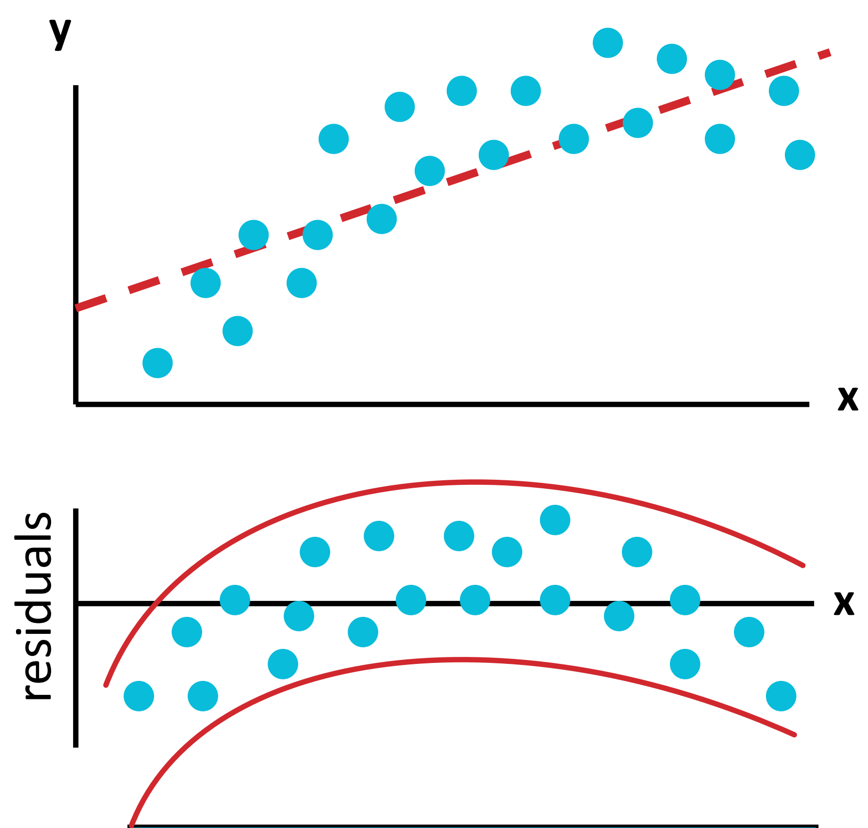
  Quadratic model

$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$$

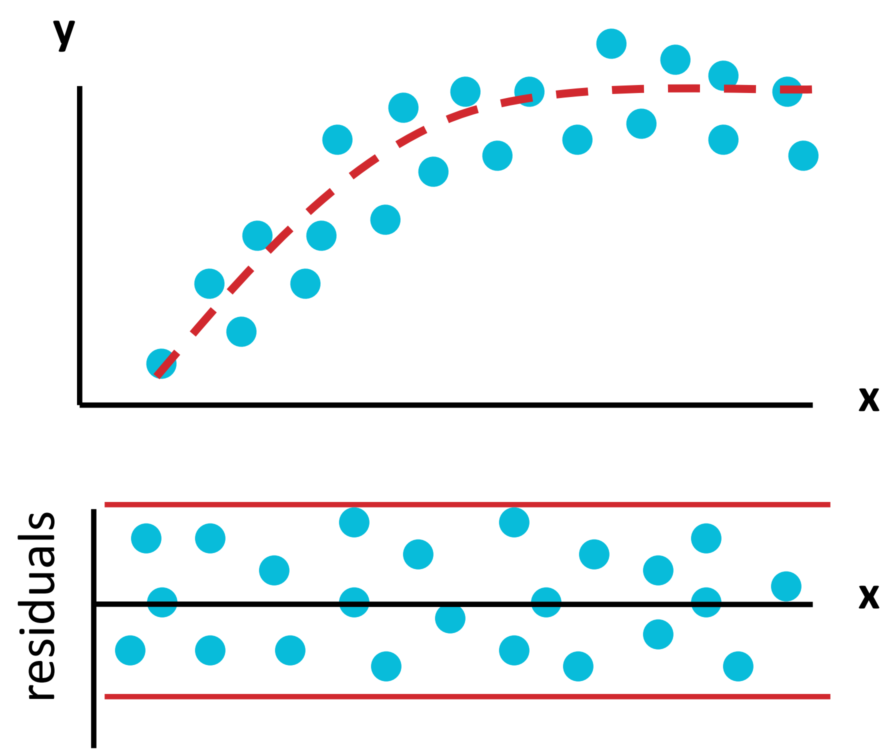
- The second independent variable is the square of the first variable.



LINEAR VS. NONLINEAR FIT



Linear fit does not give random residuals



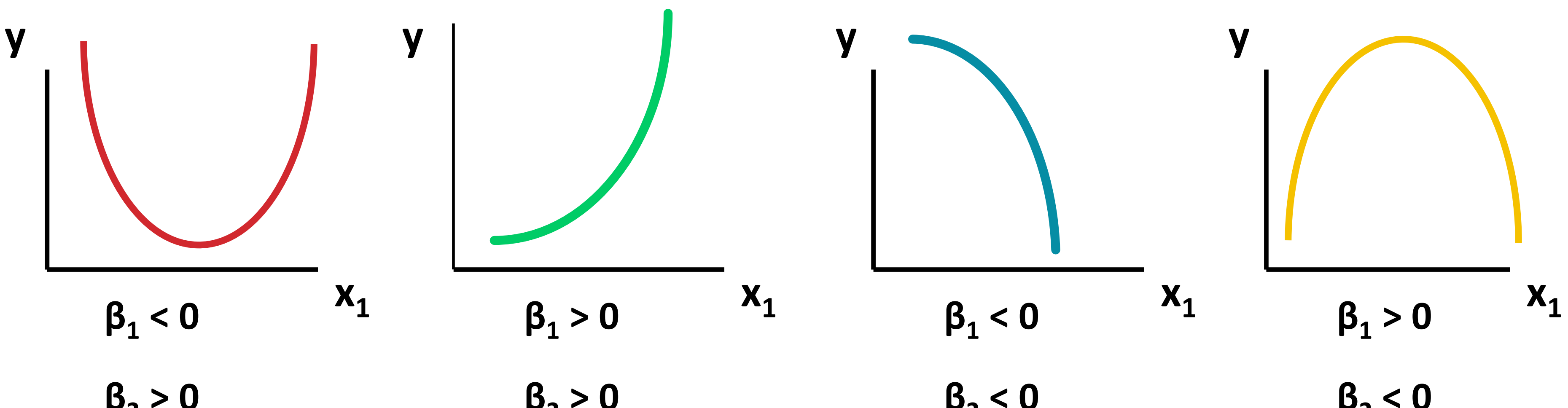
Nonlinear fit gives random residuals

QUADRATIC REGRESSION MODEL

$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$$

Quadratic models may be considered when scatter diagram takes on the following shapes:

β_1 = the coefficient of the linear term
β_2 = the coefficient of the squared term



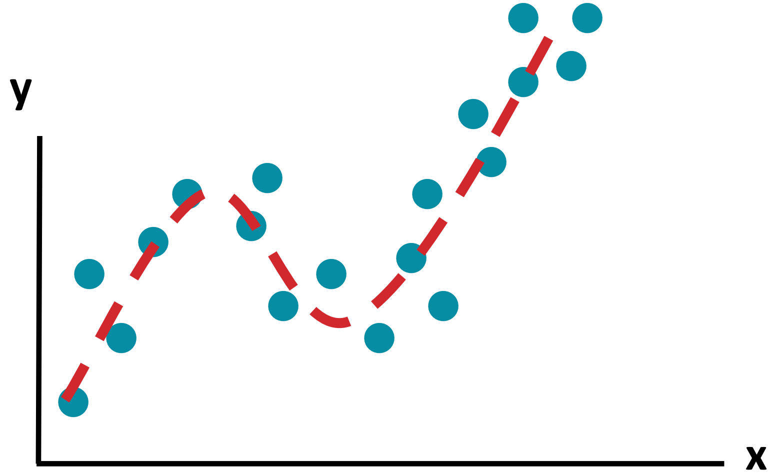
TESTING FOR SIGNIFICANCE: QUADRATIC MODEL

- Test for Overall Relationship
  - F test statistic
- Testing the Quadratic Effect
  - Compare quadratic model $y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$
  - With the linear model $y = \beta_0 + \beta_1 x_j + \varepsilon$
- Hypotheses

  H_0: $\beta_2 = 0$ (No 2nd order polynomial term)

  H_A: $\beta_2 \neq 0$ (2nd order polynomial term is needed)

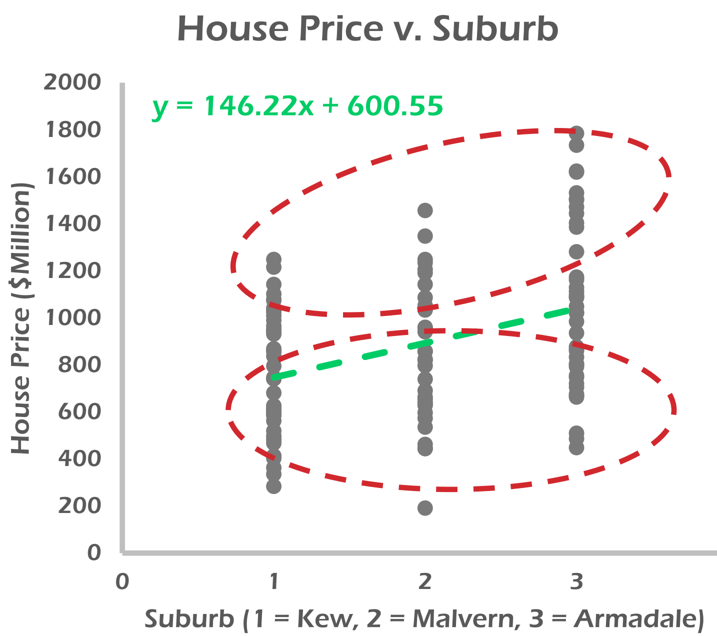
HIGHER ORDER MODELS



If $p_{(\text{order of polynomial})} = 3$ the model is a cubic form:

$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3 + \varepsilon$$

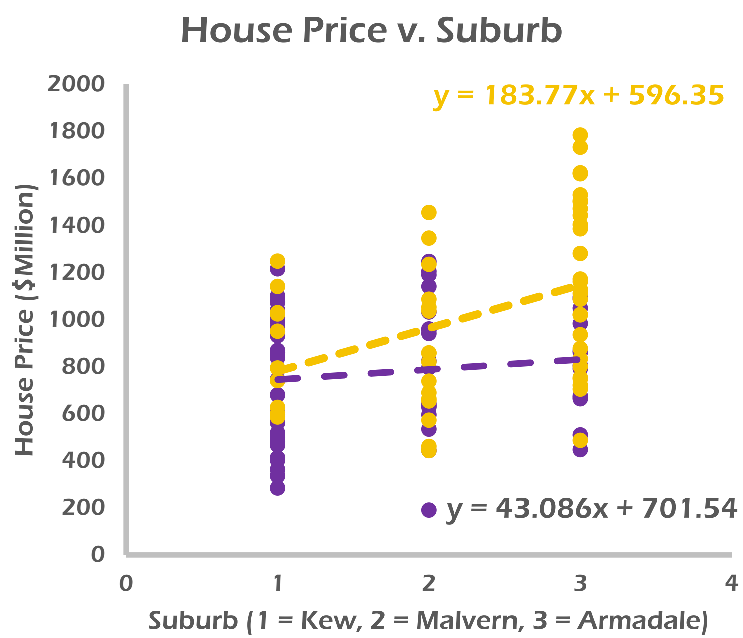
DO YOU NOTICE ANYTHING UNUSUAL?



A 3th variable may be interacting with the main independent (predictor) variable effect on the dependent variables!

DO YOU NOTICE ANYTHING UNUSUAL?

Let's introduce Gender
(Female = 0, Male = 1)



Let's introduce Style
(Traditional = 0, Modern = 1)

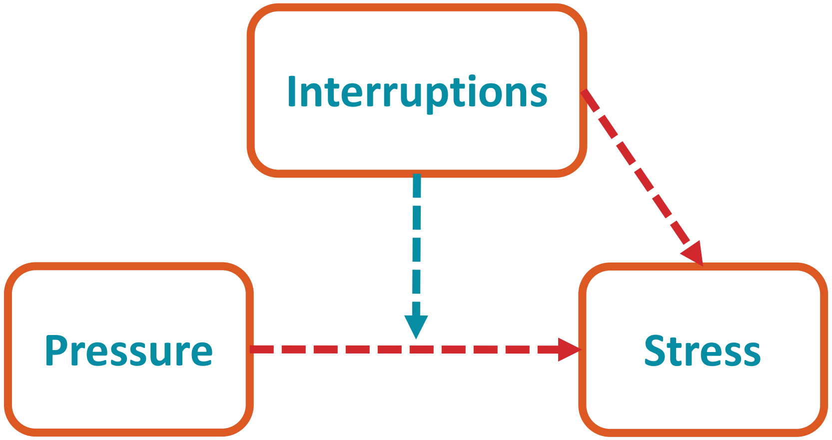
INTERACTION EFFECTS
AN EXAMPLE

BLITZ management is concerned about Job Stress among their employees and wishes to investigate key factors that may contribute to employees' stress level at work?

Research shows work stress is directly influenced by excessive work pressure. That is the higher the work pressure the higher the likelihood of employees feeling stressed at work.

INTERACTION EFFECTS
AN EXAMPLE

Employees also reported that constant interruptions at work (e.g. emails, calls, meetings etc.) adds more pressure on them and consequently may contribute to higher level of stress at work. In other words, interruptions may act together with work pressure (i.e., interact) and impose larger effect on work stress.



WHAT IS INTERACTION EFFECT?

- Hypothesise interaction between pairs of *x* variables.
  - ✓ Response to one *x* variable varies at different levels of another *x* variable.
- Contains two-way cross product terms (interaction term).

$$y = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2}_{\text{Basic Terms}} + \underbrace{\beta_3 x_1 x_2}_{\text{Interaction Term}}$$

EVALUATING PRESENCE OF INTERACTION

- Hypothesise interaction between pairs of *X* variables.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

- Hypotheses:
  - H_0: $\beta_3 = 0$ (no interaction between x_1 and x_2)
  - H_A: $\beta_3 \neq 0$ (x_1 interacts with x_2)

EFFECT OF INTERACTION

- Given: $$\boxed{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2}$$
- Without interaction term, effect of x_1 on y is measured by β_1
- With interaction term, effect of x_1 on y is measured by $\beta_1 + \beta_3 x_2$
- Effect changes as x_2 increases.

BLITZ MANAGEMENT WORK STRESS SURVEY

INTERACTION EFFECTS – SAMPLE DATA

	Dependent variable	Independent variable	Independent (interacting – moderator)	Interaction Term
ID	**Stress (y)**	**Pressure (x_1)**	**Interruptions (x_2)**	**Interaction ($x_1 * x_2$)**
1	2	2	2	**4**
2	3	3	2	**6**
3	2	3	3	**9**
4	2	3	3	**9**
5	4	3	3	**9**
6	4	4	4	**16**
7	3	4	4	**16**
8	3	3	4	**12**
9	2	1	5	**5**
10	4	5	3	**15**

y, x_1, x_2 are all measured on a 5-point Likert type scale.

multiply x_1 by x_2 to get x_1x_2, then run regression with y, x_1, x_2 , x_1x_2

INTERACTION REGRESSION OUTPUT

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.02	0.16	6.33	0.00	0.71	1.34
Pressure	0.45	0.04	12.22	0.00	0.38	0.52
Interruptions	0.18	0.04	4.56	0.00	0.11	0.26
Interaction	0.15	0.03	2.38	0.02	0.01	0.12

$b_{\text{interaction}} = 0.15$, with p-value $0.02 < 0.05$

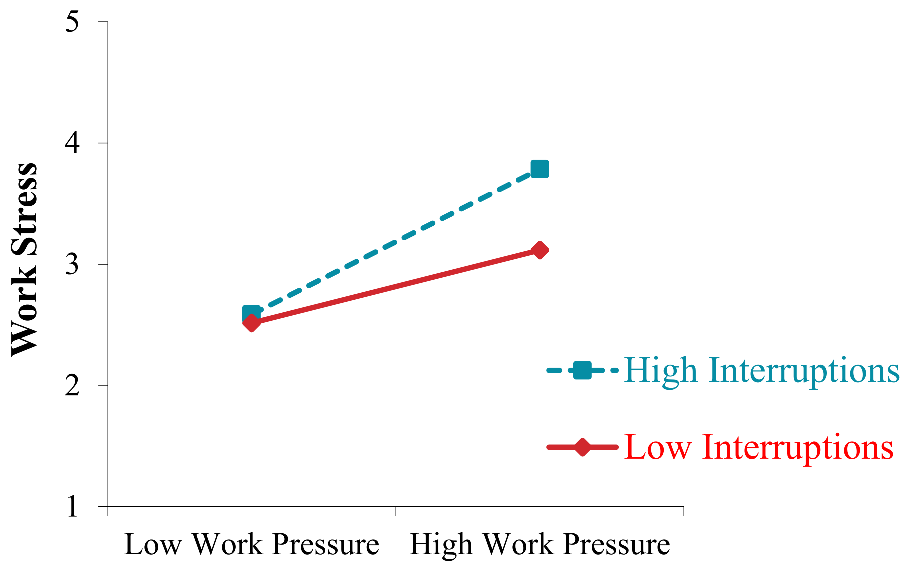
Interpretation:

We conclude that at 5% significance, interaction does exist in the regression model. In other words, Effect (slope) of work pressure on work stress does depend on interruption value.

NOTE: For interaction effect to exist, it is NOT necessary to establish a significant relationship between interacting variable and the dependent variable.

UNDERSTANDING INTERACTION EFFECTS

Interaction - Continuous



- When work pressure is low, employees experience the same level of work stress irrespective of their interruption level.
- As work pressure increases, those employees who get interrupted more often, will display higher level of work stress compared to others.
- In other words, the effect of work pressure on work stress changes (increases) as interruptions become more frequent (increases).

STEPS IN RUNNING REGRESSION ANALYSIS WITH INTERACTION TERMS

1. Decide whether or not interaction effect should be expected:
   - ✓ Prior Research and Theory
   - ✓ Scatter plots (Generally useful if intervening variable is categorical)
2. Assign variable roles:
   - ✓ Dependent variable
   - ✓ Independent variables
   - ✓ Intervening variable
3. Define interaction term by multiplying scores of independent variable by intervening variable
4. Run regression analysis by including all variables into the model
5. Check significance of the interaction term. If significant (p-value < 0.05), then interaction exists.
6. Interpret the interaction effect visually (using plotting) and in plain language