Causal (or Cause-and-Effect) type Research Question

By: Ir. Togar A. Napitupulu, MS., MSc., Ph.D

Some variables (called ***independent variables or response variables***) might be hypothesized to have and effect on, or might cause, a variable, (called the ***dependent variable or explanatory variable).***

1. The effect might be direct, as in the case of the different dosages of anti-biotic on the speed of cure of a disease.
2. Or might not have direct impact, such as in the relationship between the mileage of a car and its price.

In the case 1. above, the independent variables are fixed or determined, usually the case in experimental or quasi experimental design; while in the case 2., the independent variables are randomly distributed.

In general the statistical/mathematical model for such research question is:

$$\mathrm{Y} = f(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \ \ldots. \quad , \mathrm{X}_\mathrm{k}, \varepsilon) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

Where Y is the *effect* or the **dependent** variable, which is dependent on X_i, i=1, … k, that is the *cause*, or the **independent** variables. A particular form of the function or the relationship, and the most common one, is the following **linear** function:

$$\mathrm{Y} = \beta_0 + \beta_1\mathrm{X}_1 + \beta_2\mathrm{X}_2 + \beta_3\mathrm{X}_3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \beta_\mathrm{n}\mathrm{X}_\mathrm{k} + \varepsilon \ \ldots\ldots \text{(2)}$$

This is the ***population*** model, where the βs are the parameters that need to be estimated, usually from ***sample***. The estimated regression **line (***notice this is without* ε *term***)** is represented by:

$$\widehat{\mathrm{Y}} = \widehat{\beta}_0 + \widehat{\beta}_1\mathrm{X}_1 + \widehat{\beta}_2\mathrm{X}_2 + \widehat{\beta}_3\mathrm{X}_3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \widehat{\beta}_\mathrm{n}\mathrm{X}_\mathrm{k} \ \ldots\ldots\ldots \text{(3)}$$

where $\widehat{\beta}_\mathrm{i}$ are the estimator of β_i, *i = 1, … , k*. Other notation for $\widehat{\beta}_\mathrm{i}$ is b_i, i= 1, …. , k.

(Note: Causal here in not in the sense of actual physical cause. It could be just in term of relationship)

EXAMPLE : Possible research questions related to the above model
TELADAN : Pertanyaan penelitian yang berkaitan dengan model di atas

1. What are the main factors (or the independent variables) that affect the dependent variable (Y)? (***Note:*** *Y is usually the variable of main interest to researcher*)
   *Apa saja faktor-faktor utama (dalam hal ini sebagai variable independen) yang mempengaruhi atau sebagai penyebab dari variable dependen (Y)? (**Catatan:** Y biasanya merupakan varabel yang menjadi perhatian utama peneliti)*

   Usually researcher would start by assuming/considering (or hypothesizing) some k number of independent variables that have effects on the dependent variable (see equation 2). The objective then is to test hypothesis on all parameters corresponding to each independent variables whether they are equal to zero or not, based on sample data, i.e.,

   $H_0: \beta_i = 0$
   $H_1: \beta_i \neq 0$

   for each *i = 1, … k*. If we fail to reject H_0 (or it is accepted) for a particular *i*, then we can conclude that the corresponding X_i does not affect (or cause) Y. Or in other words, X_i should not be in the model or equation (2). Otherwise, X_i would be one of the independent variable that causes Y. How do we decide whether to reject or accept H_0? The easiest way is to look at the ***p-value*** (*p-value* is generated by almost all statistical packages – SPSS for example give it a name P or Sig.). The *p-value* actually is the committed error as a result of rejecting a true Null Hypothesis (H_0). Therefore, if *p-value* is **less than** the ***allowable error (significant level) α,*** then we might as well reject H_0, since our committed error by rejecting it is less than the allowable error, α. Most common α is 5%; but for management, social and behavioral sciences it could go up to 10%.

   Biasanya pada pertanyaan penelitian seperti ini, peneliti akan memulai dengan mengasumsikan (tentu berdasarkan teori) atau menghipotesakan bahwa ada sebanyak k variable independen yang mempengaruhi atau berdampak terhadap variable dependen (lihat persamaan 2). Tujuan penelitian tentu adalah menguji hipotesis terkait masing-masing parameter dari tiap-tiap variable independen, apakah masing-masing sama dengan nol atau tidak, berdasarkan data dari sampel, yaitu,

   $H_0: \beta_i = 0$
   $H_1: \beta_i \neq 0$

   Untuk tiap-tiap i = 1, …. , k. Bila kita gagal menolak Ho (atau dengan kata lain Ho diterima) untuk i tertentu, maka kita dapat menyimpulkan bahwa variable X_i tidak mempengaruhi atau tidak menjadi penyebab dependen variable Y. Atau dengan kata lain, X_i tidak termasuk dalam persamaan 2. Sebaliknya, X_i adalah salah satu factor penyebab atau yang mempengaruhi

*variable Y. Bagaimana caranya menguji apakah menolak atau menerima Ho? Cara yang paling gampang adalah dengan melihat **p-value** (p-value dihasilkan oleh hampir semua paket perangkat lunak statistika – SPSS misalnya menamainya P atau Sig.). p-value sebenarnya adalah kesalahan yang timbul sebagai akibat penolakan terhadap Ho yang benar. Oleh karena itu jika p-value lebih kecil dari error yang diizinkan (**significant level) α**, maka tidak ada masalah bila kita menolak Ho karena kesalahan yang kita perbuat sebagai akibat penolakan tersebut lebih kecil dari α, yaitu kesalahan yang diizinkan. Besarnya α yang umum adalah 5 %, sekalipun tingkat significansi 10% masih ditolerir terutama untuk bidang ilmu sosial, manajemen dan ilmu-ilmu perilaku.*

2. Once a particular β is proven significant (i.e., by rejecting H_0), or that the corresponding X is significantly affecting Y, researcher might be further interested in the **magnitude** of the β itself, estimated by $\hat{\beta}$. For example, we might want to compare the magnitude of the impact among those independent variables on the dependent variable. Or for that matter, we might want to compare the β corresponding to the same particular independent variable but with different dependent variable (in this case we are comparing two regression equations for example).

   *Begitu β tertentu telah dibuktikan secara signifikan tidak sama dengan nol (yaitu, dengan menolak H_0), atau bahwa variable X nya secara signifikan mempengaruhi Y, peneliti selanjutnya mungkin tertarik dengan **besarnya** pengaruh tersebut yang digambarkan oleh β, dengan $\hat{\beta}$ sebagai penduganya. Sebagai contoh, kita mungkin ingin membandingkan besarnya dampak diantara variable independen terhadap variable dependennya. Atau barangkali kita ingin membandingkan besar pengaruh variable independen yang sama namun terhadap variable dependen yang berbeda (dalam hal ini kita membandingkan dua persamaan regressi misalnya).*

   In finance for example, we are familiar with the so called "the beta", that is the coefficient of regression of a particular stock return against a particular market index (for example TELKOM stock return against Index Harga Saham Gabungan (IHSG); or GOOGLE stock return against Standard and Poor's (S&P) 500 index) as a measure of a systematic risk of a stock or an asset. Fund managers might classify stocks with $\hat{\beta} > 1$ and those with $\hat{\beta} < 1$. When the market is good, indicated by increasing IHSG, then it is better to put the fund in stocks with $\hat{\beta} > 1$; otherwise, should be put in stocks with $\hat{\beta} < 1$.

   Dalam bidang keuangan misalnya, kita familiar dengan konsep "the beta", yaitu coefisien regressi dari saham tertentu terhadap indeks saham gabungan (misalnya, hubungan regressi antara harga saham TELKOM terhadap Indeks Harga Saham Gabungan (IHSG); atau harga saham GOOGLE terhadap indeks S&P 500) sebagai ukuran resiko sistematik dari suatu saham atau asset. Fund managers biasanya mengklasifikasikan stocks dalam dua keompok, yaitu, kelompok dengan $\hat{\beta} > 1$ dan kelompok dengan $\hat{\beta} < 1$. Apabila kondisi pasar baik, yang

diindikasikan dengan peningkatan IHSG misalnya, maka Fund manager sebaiknya menaruh dananya di kelompok stocks dengan $\hat{\beta} > 1$*; sebaliknya bila keadaan pasar kurang bagus, maka sebaiknya dana ditaruh di kelompok stocks dengan* $\hat{\beta} < 1$.

In general, the magnitude of β, measures the sensitivity of the dependent variable against the independent variable. It measures the % changes in Y as result of a 1% change in X. The beta of a regression between quantity demanded against the price of a product for example, is an important measure of sensitivity called elasticity in economics.

Secara umum, besarnya β *mengukur sensitifitas dari variable dependen terhadap perubahan pada variable independen. Dia mengukur % perubahan pada Y sebagai akibat perubahan X sebesar 1 %. Sebagai contoh, beta dari regressi antara jumlah permintaan akan suatu produk tertentu terhadap perubahan harga merupakan ukuran penting yang disebut elastisitas permintaan terhadap harga dalam bidang ekonomi.*

In some cases, the βs might not be relevant at all, such as when the independent variables and the dependent variable are concept or latent variables. For example, consider a regression between loyalty and satisfaction. Percent change in satisfaction might not be meaningful at all; similarly it might be not easy to imagine % increase in Loyalty. However, comparing among the βs within one model even though the variables are latent variables, might still be relevant, i.e., they enable the researchers to compare the ranking (order) of the impact caused by the independent variables. Almost all statistical packages provide feature called "*standardized*". This feature allows you to have estimator of the β in a standardized unit, meaning that their magnitudes are comparable regardless of the units used for measuring the independent variables (the X_is).

Dalam beberapa kasus, βs *mungkin tidak terlalu relevan, seperti untuk kasus dimana variable independen dan dependen nya merupakan variable konsep atau laten. Sebagai contoh, perhatikanlah regressi antara loyalitas dengan kepuasan. Persentasi perubahan untuk variable kepuasan mungkin tidak terlalu bermakna; demikian pula dengan loyalitas, tampaknya tidak terlalu mudah membayangkan atau mencari arti dari persentase kenaikan pada loyalitas. Akan tetapi, membandingkan* βs *dalam satu model sekalipun variable dependen nya adalah laten, bisa saja masih relevan, yaitu, peneliti dimungkinkan membandingkan urutan dampak dari antara independen variable terhadap dependen variable. Hampir semua paket perangkat lunak statistika memiliki feature yang memungkinkan peneliti mendapatkan estimator dari* β *yang "terstandardisasi". Ini berarti besaran dari penduga dari beta dapat dibandingkan satu sama lainnya sekalipun unit pengukuran masing-masing variable X tidak sama.*

3. Regression analysis also generates the so called "***coefficient of determination***" with common symbol $\mathbf{R^2}$. We know that the following formula hold for regression:

$$\sum_{i=1}^{n}(Y-\bar{Y})^2 = \sum_{i=1}^{n}\left(Y-\hat{Y}\right)^2 + \sum_{i=1}^{n}\left(\hat{Y}-\bar{Y}\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Where the term in the left hand side of the equation is the "total sum of squares of deviation of observations Y_i around Y average" with symbol SST; the first term on the right side of equation (4) is "the sum of squares of the deviation of observations Y_i from the regression estimates" = SSR; and the second term in the right side of equation (4) is " the sum of squares of deviation of the regression estimates and the average of Y" = SSE. Equation (4) then can be rewritten as follows:

$$SST = SSR + SSE \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

and R^2 is now defined as:

$$\mathbf{R^2} = \frac{SSR}{SST} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

That is, that $\mathbf{R^2}$ provide measurement of the variation in Y, the dependent variable, explained by the regression. Or in other words, $\mathbf{R^2}$ reflects the ability of the regression to explain Y, the dependent variable.

Now, with the explanation above, $\mathbf{R^2}$ can be used to answer research question such as reliability of using β-coefficient for example as a measurement of systematic risk in finance (see example 2 above). For example, $\mathbf{R^2}$ = 0.8 can be interpreted that we can rely 80% upon the β as measurement of the change in our stock price as a result of a 1% change in IHSG or S&P 500 index.

Another research question that can be addressed using $\mathbf{R^2}$ is when we want to know the contribution of independent variable(s) in trying to explain the variation in the dependent variable. For instance, suppose competitiveness can be measured, may be in terms of financial measures, or any of those competitiveness measurement based on Norton's balanced score card for that matter; and we want to know to what extend does IS/IT strategy (a variable or variables that also can be measured or observed) contributes to competitiveness of the company. This extend then can be derived from R^2 by regressing competitiveness (Y) against IS/IT strategy (X) dependent of which measures are being used to represent Y. It is possible for example that for Restaurant and Hotels industries the R^2 with respect to "financial" measure as our Y is small but for "customer relation" as our Y might be a lot bigger.

4. Regression also can be used to predict the value of Y given the value of Xs. In such case we are actually building a regression model or a "machine", a learning machine by estimating the parameters of the regression. During the estimation of the parameters of the regression, in the jargon of data mining, we say that the machine is learning from the data. It is sometimes called "supervised learning", because the parameters are forced (supervised) to fit the pairs of Y and the Xs.

Consider for Example ones want to build a "machine" (a regression) that is able to distinguish non-performing loan vs. well-performing loan customers. The Y then would be loan performance of the customers having values of 1 for well-performing borrower and 0 for non-performing borrower. The Xs would be the variables or characteristics of the borrowers that we think might be related to or to cause performance of the borrowers. Notice however, because Y is binary (0,1), it violates the normality requirements of the traditional Ordinary Least Square Method of estimation. To resolve this, we take the regression of the logarithm (base 2) of the "odds ratio" against the Xs, following this logistic model:

$$\ln(\text{odds ratio}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_n X_{ki} + \varepsilon_i \quad \ldots \quad (7)$$

Where

$$\text{odds ratio} = \frac{\text{probability of an event of interest}}{1 - \text{probability of an event of interst}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

Suppose an event of interest is well-performing loan. Probability of a well-performing loan (or the probability of an event of interest) then is the frequency of performing loan in the sample. The estimated regression equation is:

$$\ln(\widehat{\text{odds ratio}}) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots + \hat{\beta}_n X_k \quad \ldots \quad (9)$$

Odds ratio then (the estimated one) is

$$\text{odds ratio} = e^{(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \ldots\ldots\ldots\ldots\ldots + \hat{\beta}_n X_k)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

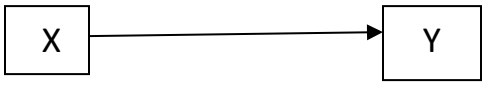
Having known the estimated odds ratio, we can now substitute it back to equation (8) to get the (estimated) probability of an event of interest as follows:

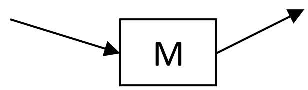
$$\text{Probability of an event of interest} = \frac{\text{estimated odds ratio}}{1 + \text{estimated odds ratio}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

The learning process stops right at equation (9) the "machine". Next, this machine can now be used to predict probability of a new particular customer whether he is going to be a well performing borrower using equation (11).

This is actually a special case of "Discriminant analysis", where the possible value of Y not only 1 or 0; but could be more than two categories.

5. In research we are familiar with the so called "intervening" or "mediating" variable. Modeling of such relationship usually depicted as follows:





Variable M is the mediating variable. The model assumes that there is a direct effect from X to Y. However, we also hypothesized that there is a path that first goes through M and then from M on to Y. Such structure is modeled in regression using two equations as follows:

$$Y = \beta_{10} + \beta_{11}X + \beta_{12}M + \varepsilon_1 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \quad (1)$$

$$M = \beta_{20} + \beta_{21}X + \varepsilon_2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (2)$$

The direct effect from X to Y occurs when we reject H_0 in the following hypothesis testing

H_0: $\beta_{11} = 0$ versus

H_1: $\beta_{11} \neq 0$

To test the effect of X on Y through the mediating variable M, is equivalent to testing the following hypothesis:

H_0: $\beta_{21} = 0$ versus

H_1: $\beta_{21} \neq 0$

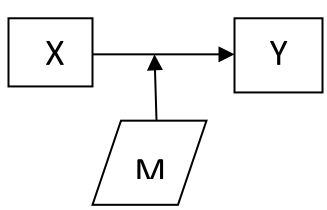
And

H_0: $\beta_{12} = 0$ versus

H_1: $\beta_{12} \neq 0$

Where both H_0s in the two sets of hypothesis is rejected. These tests can be done using SPSS-LISREL or SPSS-AMOS.

6. Another common research question is when there is a causal relationship between two variables and we are hypothesizing that the relationship is conditional on another variable usually called “moderating” variable. Such model is depicted in the following figure:



Suppose M is a binary variable indicating two category, eg., Male or Female. Then statistically, the above model can be formulated by introducing dummy variable *D* as follows:

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 XD + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (4)$$

Where $D = \begin{cases} 1, & if\ the\ observation\ is\ Male \\ 0, & if\ the\ observation\ is\ Female \end{cases}$

To test whether there are actually two parallel regression, one for Male and one for Female, we can test whether $H_0: \beta_2 = 0$ or not, i.e., $H_1: \beta_2 \neq 0$. Again if we reject H_0 then the effect of moderating variable is significant, meaning that there should be two regressions, one for Male and one for Female. Once it is proven that the moderating variable is significant, we can further test whether the slope of the two regression is the same or not, i.e., they are parallel or not. This is can be tested by testing $H_0: \beta_3 = 0$ versus $H_1: \beta_3 \neq 0$. If H_1 is accepted, then there is an interaction effect between gender and X on Y. Meaning that the slope of the regression of X on Y for both gender are different. Or they are not parallel. The data for variable *XD (the interaction)* is generated from multiplying variable *X* and *D.*

The following are the detail cases to be considered:

Case #1: $\beta_2 \neq 0, \beta_1 \neq 0, \beta_3 \neq 0$

When $\beta_2 \neq 0$, it means the impact of moderating variable, represented by dummy *D*, is significant. It means that equation (1), for male, i.e., *D=1,* become:

$$Y = \beta_0 + \beta_1 X + \beta_2 + \beta_3 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (2)$$

Or

$$Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (2 - male)$$

And for female (*D=0),*

$$Y = \beta_0 + \beta_1 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (2 - female)$$

Or the slope for male is different by β_3 and the intercept is different by β_2 from female.

Case #2: $\beta_2 \neq 0, \beta_1 \neq 0, \beta_3 = 0$

The impact of moderating variable and the X is significant, while the impact of interaction *XD* is not significant. Equation (1) then for **male** become:

$$Y = \beta_0 + \beta_1 X + \beta_2 + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Or

$$Y = (\beta_0 + \beta_2) + \beta_1 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (3 - male)$$

And for female become:

$$Y = \beta_0 + \beta_1 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (3 - female)$$

So, the slopes are the same while the intercept are different by β_2.

Case #3: $\boldsymbol{\beta_2 \neq 0, \beta_1 = 0, \beta_3 \neq 0}$

Equation (1) for male (*D*=1) become :

$$Y = (\beta_0 + \beta_2) + \beta_3 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (4 - male)$$

And for female (*D=0*),

$$Y = \beta_0 + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (4 - female)$$

Case #4: $\boldsymbol{\beta_2 = 0, \beta_1 \neq 0, \beta_3 \neq 0}$

Equation (1) for male become :

$$Y = \beta_0 + (\beta_1 + \beta_3) X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (5 - male)$$

And for female become:

$$Y = \beta_0 + \beta_1 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (5 - female)$$

Case #5: $\boldsymbol{\beta_2 = 0, \beta_1 \neq 0, \beta_3 = 0}$

Equation (1) for male become :

$$Y = \beta_0 + \beta_1 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (6 - male)$$

And for female become:

$$Y = \beta_0 + \beta_1 X + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (6 - female)$$

Case #6: $\boldsymbol{\beta_2 = 0, \beta_1 = 0, \beta_3 = 0}$

Equation (1) for male (*D=1*) become:

$$Y = \beta_0 + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (7 - male)$$

And for female (*D=0*):

$$Y = \beta_0 + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (7 - female)$$

In the case where you have *n* category of moderating variable, then you would need *(n-1)* dummy variables to represent them.

7. ONE COMMON MISS-USED OF F-test (ANOVA) IN REGRESSION: Testing Simultaneous Effect of Independent Variables

In regression analysis of the following model (without loss of generality, let's have three independent variables):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

one might wants to test whether the three independent variables have an impact simultaneously on the dependent variable Y. It is common to have this tested (mistakenly) by using the ***F-test*** from the ANOVA of the regression, or equivalently by looking at corresponding *p-value*. It is mistaken because the hypothesis corresponding to such *F-test* is:

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ against
H_1: At least one of the betas is not equal to zero

So for example if we reject H_0, it only means that one or more of the β_i is not equal to zero and to know which one that is not equal to zero, can be tested partially using ***t-test***.

How then we can test hypothesis to find out the effect of all independent variables simultaneously? We approach this by introducing a new variable into equation (1), that is, multiplication of all independent variables as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 X_3 + \varepsilon \dots\dots\dots\dots\dots\dots\dots\dots (2)$$

Now, to test such hypothesis then is equivalent to testing

H_0: $\beta_3 = 0$ against
H_1: $\beta_3 \neq 0$

If from the data there enough evidence to reject H_0 then it can be concluded that the three independent variables interact (*hence the term to test the "interaction" of the independent variables)* or simultaneously has an impact on the dependent variable Y. Researchers might want to test the impact of interaction of pairs of the independent variables. In such case, we would have to introduce three more independent variables into equation (2) as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3 + \varepsilon \ldots (3)$$

And testing the impact of the interaction would be testing the corresponding betas, i.e., β_i, *i =4,5,6,7.*